Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# On the limits of detecting introgression from early modern human dispersals

**Oscar Lao**

oscar.lao@csic.es

Institut de Biologia Evolutiva

**Jose Serradell Noguera**

Institut de Biologia Evolutiva https://orcid.org/0000-0001-6551-1432

**Alba Nieto**

**Jaume Bertranpetit**

CSIC-Universitat Pompeu Fabra https://orcid.org/0000-0003-0100-0590

**Article**

**Keywords:**

**Posted Date:** November 16th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11376410/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** There is **NO** Competing Interest.

# On the limits of detecting introgression from early modern human dispersals

Serradell, Jose M[1*]., Nieto, Alba[2], Bertranpetit, Jaume[1], Lao, Oscar[1*].

[1] Institut de Biologia Evolutiva, IBE, (CSIC-Universitat PompeuFabra), Barcelona, Spain

[2] Center for Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS), RIKEN,2-1 Hirosawa, Wako, Saitama 351-0198, Japan

*josemiguel.serradell@ibe.upf-csic.es

*oscar.lao@csic.es

## Abstract

The expansion of *Homo sapiens* out of Africa 50–60 thousand years ago (kya) gave rise to all ancestral lineages outside Africa. However, fossil evidence from the Levant, East Asia, and Southeast Asia dated between ~70 and 210 kya raises questions about an earlier Out-of-Africa (eOOA) expansion. Genomic analyses have failed to detect such expansion in present-day populations, supporting complete replacement by the later expansion. An alternative hypothesis is that introgression from eOOA populations is intrinsically difficult to detect. We investigate the demographic conditions under which introgression from an eOOA population becomes detectable in genomic data. We examined the distribution of genomic fragments under coalescent processes. While long fragments associated with deep Times to the Most Recent Common Ancestor are expected under introgression, we find that under plausible demographic scenarios such signals become indistinguishable from background variation. We then evaluated realistic demographic models using an ABC-DL framework applied to real genomes. Models with and without eOOA introgression were only weakly distinguishable, with models lacking eOOA introgression providing the best fit to the observed data. Our findings help reconcile the apparent discrepancy between fossil and genomes and highlight the need for methods specifically designed to detect introgression between closely related populations.

# Introduction

The dispersal of *Homo sapiens* from Africa represents one of the defining demographic events in human evolutionary history. Genetic evidence indicates that the majority of ancestry in present-day non-African populations derives from a major expansion that has been estimated to have occurred approximately 50–60 thousand years ago (kya) from the African continent[1–3].

During this expansion, modern humans encountered and admixed with other hominin groups across Eurasia, including Neanderthals and Denisovans. These interactions left clear genomic signatures of introgression that remain detectable in present-day populations and have become a central component of models of recent human evolution[3–7].

Despite this well-supported genomic framework, the archaeological and fossil records increasingly suggest that the dispersal history of modern humans may have been more complex. Fossil discoveries from several regions indicate that anatomically modern humans were present outside Africa substantially earlier than the Late Pleistocene expansion associated with present-day populations. In the Levant, fossils from Skhül and Qafzeh caves document the presence of *Homo sapiens* between approximately 90 and 130 kya[8–10]. Even earlier evidence has been reported from Misliya Cave in Israel, where a maxilla attributed to modern humans has been dated to approximately 177–194 kya, suggesting that members of the *Homo sapiens* lineage had already dispersed beyond Africa during the Middle Pleistocene[11]. Additional fossil and archaeological evidence from East and Southeast Asia further suggest that modern humans may have reached parts of Eurasia well before the main Late Pleistocene dispersal[10,12–14].

These findings have led to the hypothesis that earlier dispersals of modern humans out of Africa—often referred to as early Out-of-Africa (eOOA) events—may have occurred repeatedly during the last several hundred thousand years[15,16]. Under this scenario, populations of *Homo sapiens* expanded intermittently into Eurasia during periods of favorable climatic conditions and may have established local populations that later disappeared or were replaced by subsequent expansions. An

important question arising from this hypothesis is whether populations associated with these early dispersals interacted genetically with the populations involved in the later and more successful expansion around 50–60 kya.

One possible explanation reconciling fossil evidence and genetics is that early dispersal populations did not contribute genetically to later human populations and were ultimately replaced by the later expansion[17]. An alternative possibility, however, is that genetic contributions from such populations exist but remain undetectable with current genomic approaches. Whereas the first explanation fits Ockham's razor expectations, the second is in line with the experience of metrics for detecting admixture and archaic introgression, which tend to be evolutionary model specific[18]. In particular, detecting introgression in genomic data depends on several demographic and evolutionary factors, mainly including the divergence time between the source and recipient populations and the proportion of admixture.

Early dispersals of modern humans present a particularly challenging case for genomic detection because the divergence between the populations involved may have been relatively shallow. If an early dispersal occurred around ~200 kya and the later expansion around ~60 kya, the genetic differentiation between these populations would be far smaller than that separating modern humans from Neanderthals or Denisovans. Under such circumstances, introgressed segments may not display sufficiently elevated divergence to distinguish them from background variation, potentially rendering them indistinguishable from ordinary within-lineage variation.

This raises a key methodological question for interpreting human evolutionary history: does the absence of detectable genomic signals necessarily imply that early dispersals did not contribute genetically to later populations? More broadly, under what demographic conditions would introgression from early modern human populations become detectable in genomic data?

Previous studies have shown that introgression from unsampled or "ghost" populations can sometimes be detected indirectly through genomic patterns such

as unusually deep genealogical segments or unexpected allele-sharing relationships[19–24] .

Here we use simulation-based approaches to define the demographic conditions under which introgression from closely related populations becomes distinguishable from background variation in genomic data. We model scenarios spanning a wide range of divergence times and introgression proportions and evaluate their detectability using complementary theoretical and ABC-based analyses. We show that introgression from early modern human populations falls largely outside the detectable region of parameter space, providing an alternative explanation for the apparent discrepancy between the archaeological and genomic records beyond a single OOA event for all modern human lineages.

## Results

### *Defining the limits of introgression detectability using simulated data*

We simulated a simplified human-like demographic model based on Gravel et al.[25], consisting of four populations, including a source population contributing introgressed ancestry to an ancestral Eurasian population (Figure 1) and assuming a 29 years/generation (see Materials and Methods). The simplified framework excludes migration and population substructure to isolate the effects of divergence time and introgression proportion on genomic detectability.

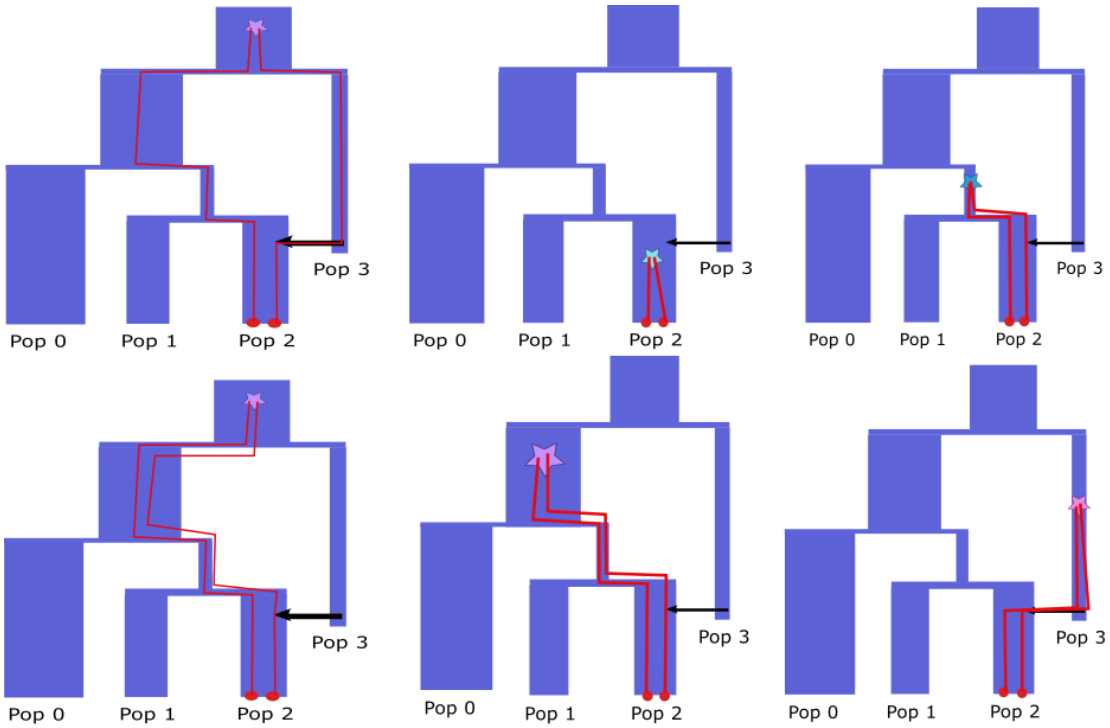



*Figure 1: Possible coalescence points of two genomic fragments extracted from population 2 in a simple introgression model. To detect introgression without a reference ancient genome we are interested in the events where fragments coalesce before the split between source and sink populations as shown in the top left figure. Bottom left figure represents background variation that would generate deeply divergent fragments that confound our ability to detect introgressed fragments.*

Signals of introgression, under a coalescent framework, can be explained as larger genomic fragments that present a deeper Time of Most Recent Common Ancestor (TMRCA) between two sampled genomes than expected just by recombination[26–28]. Under this framework, we checked how the distribution of the length and TMRCA of fragments of two sampled genomes in the target population changes when comparing scenarios of different introgression rates and divergence times of the archaic group from the current populations against a non-introgressed scenario (Figure 2).

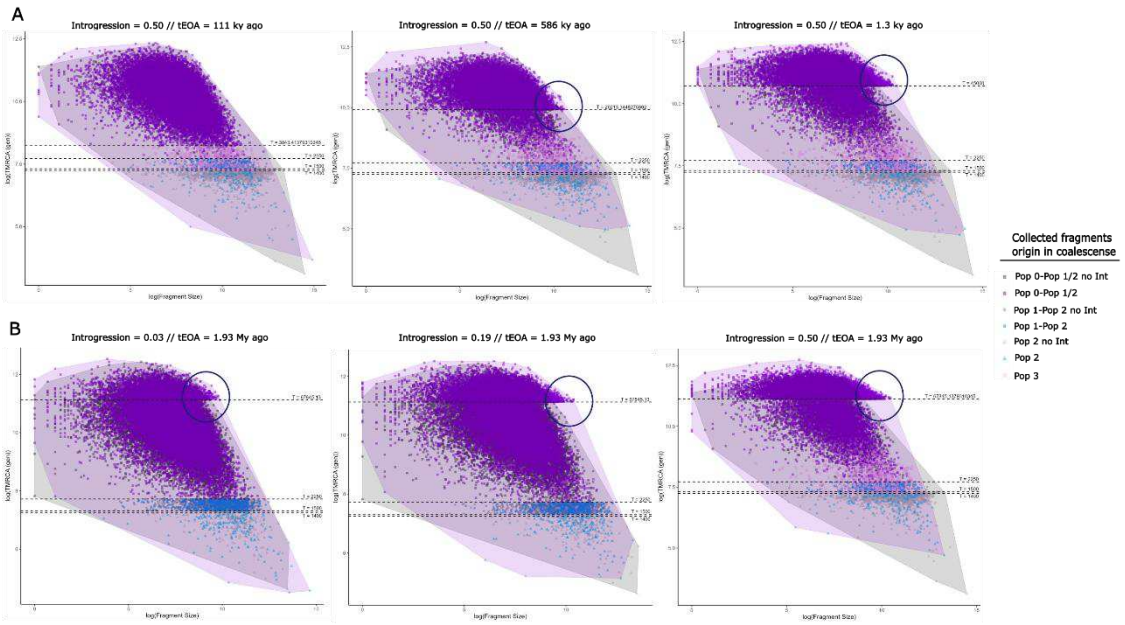



*Figure 2: Comparing distribution of fragments depending on the fragment length and the TMRCA of each fragment in logarithmic scale with recombination rate = 1e-8. A. Distributions of fixed introgression rate = 0.5 and non-introgression models. From left to right, the TMRCA between the source and sink population are 111 kya (left), 586 kya (middle), 1.3 Mya (right). Circled we see the excess of long-deep fragments that can be identified as introgressed. As the divergence time is closer the signal becomes less evident, disappearing at divergence times close to 100K. B. Distributions with fixed TMRCA of 1.9 Mya and introgression rates (left to right) of 0.03, 0.19 and 0.5. In all three scenarios we observe the excess of deep and long fragments even in low introgression scenarios. Divergence time acts as a driver of introgression detection more than introgression rate.*

First, in order to quantify the effect of the divergence time between the archaic group and the current populations on the possibility of identifying an eOOA model versus a null model of non-eOOA admixture, we considered the most favorable introgression scenario, where the amount of introgression is 0.5, meaning that 50% of the source ("ghost archaic") genome is present in the receiving population at time 0, and simulated demographic scenarios where the eOOA divergence time is in between 90k-2M years ago. Out of the 3000 simulations, we observe that simulations under the eOOA model tend to show a higher proportion of larger fragments at deeper divergence times compared to the non-introgressed model

across all different recombination rates tested (Figure 2.A & Sup Figure 1). Next, we aimed to know how the change in introgression rate affected our ability to observe the signal. For this, we fixed all simulation to a divergence time of 2 My (assuming a super deep divergence between archaic genomes and current day populations) and allow simulation to change between 0.01 to 0.5 introgression rate. We can observe the excess of larger fragments similar even when the introgression rate is very small (Figure 2.B & Sup Figure 2). As the amount of recombination could mask our ability to detect bigger fragments that could be identified as introgressed, robustness against the effects of recombination was tested as we replicated the analysis with different levels of uniform recombination.

We assessed whether divergence time and introgression proportion influence our ability to distinguish introgressed from non-introgressed fragment distributions. For each parameter combination, we calculated a weighted Jaccard distance between 100 introgressed simulations and the non-introgressed model focusing on the fragments that coalesced at deeper times (Figure 1), with higher values indicating greater dissimilarity. TMRCA showed a positive correlation with Jaccard distance at all three introgression proportions of 0.01 ($R = 0.475$, $p < 0.05$), 0.05 ($R = 0.944$, $p < 0.05$) and 0.5 ($R = 0.996$, $p < 0.05$) (Figure 3). On the other hand, introgression proportion was positively correlated with the weighted Jaccard distance at fixed TMRCA of: 300 kya ($R = 0.174$, $p > 0.05$) and 2 Mya ($R = 0.987$, $p < 0.05$) but not at TMRCA of 90 kya (R = -0.094, $p > 0.05$) (Figure 3).

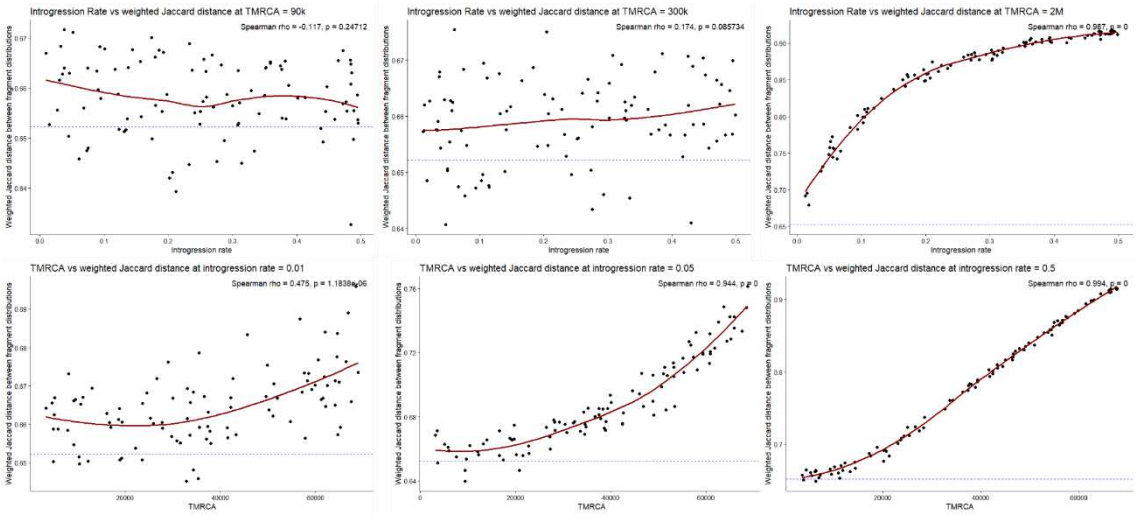



*Figure 3: Spearman correlations of the weighted Jaccard distance of the distribution of fragments between introgression and non-introgression demographic scenarios. Top row: simulations where the introgression rate is fixed and the TMRCA is variable. We observe a significant (p <0.05) positive correlation between TMRCA and weighted Jaccard distance at all different introgression rates tested. Bot row: simulations where the TMRCA is fixed and the introgression rate is variable. We observe a significant (p <0.05) positive correlation between introgression rate and weighted Jaccard distance at deeper TMRCAs while there is no correlation at low TMRCA. Blue dashed line corresponds to a weighted Jaccard distance between 2 distributions without introgression as a basal representation (wJaccard = 0.6522).*

We then model the effect of both variables and their interaction as a Beta regression. Beta regression revealed that the weighted Jaccard distance between introgressed and non-introgressed fragment distributions increased significantly with TMRCA (Eq.1). The significant positive interaction between introgression rate and TMRCA indicated that the effect of introgression became stronger as the divergence time increased (β = 5.104e-05, $p$ < 2e-16). Thus, introgressed genomic fragments were more clearly distinguishable from the non-introgressed background under scenarios combining deeper divergence times with larger introgression proportions. The model showed strong explanatory power (pseudo-$R^2$ = 0.9388) (Figure 4).

$$logit(E[DJ]) = 0.6190 - 0.4423Ir + 3.338 \times 10^{-6}T + 5.104 \times 10^{-5}IrT \quad \text{(Eq. 1)}$$

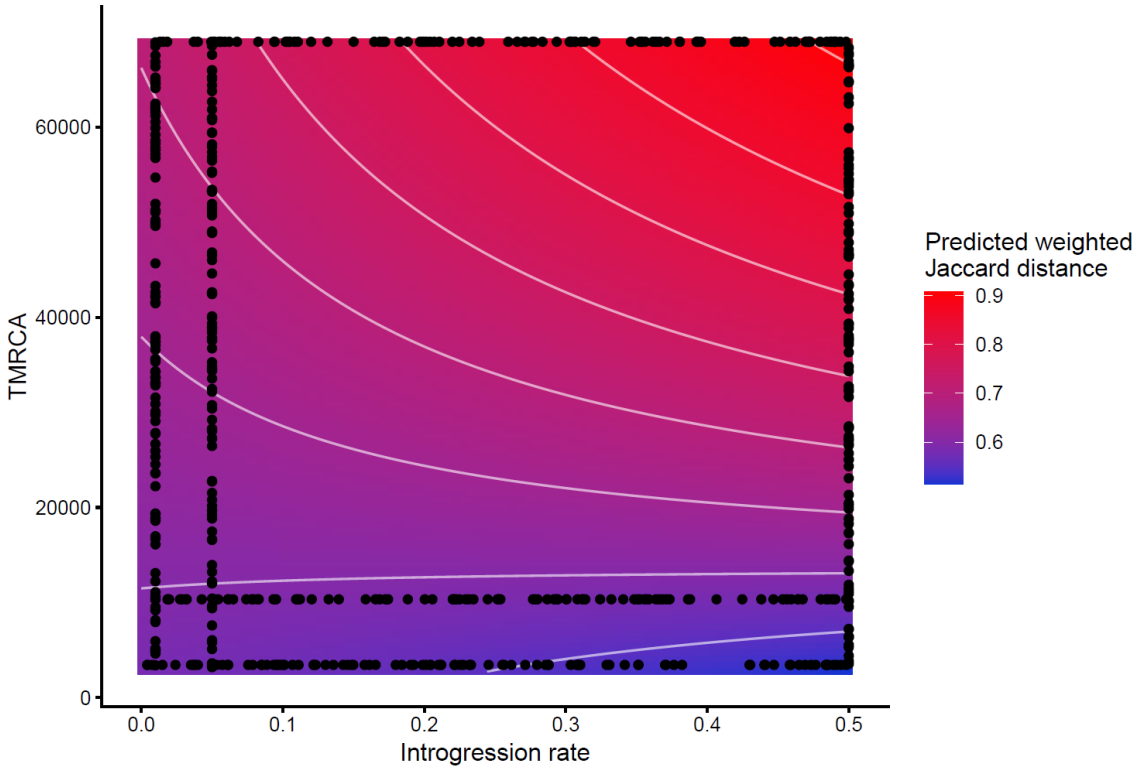



*Figure 4: Beta regression model prediction of the weighted Jaccard distance between introgressed and non-introgressed simulations. The distance between distributions mainly depends on the interaction between TMRCA and the introgression rate with the effect of introgression increase significantly with the TMRCA.*

Next, we evaluated whether introgression could be detected under parameters of divergence time and amount of introgression that more closely resemble a realistic scenario for human populations. We defined prior distributions for the amount of introgression (0 - 0.20), the TMRCA between the source and target populations (90Ky - 700Ky) and the recombination rate (2.5e-9 - 2.5e-8), while keeping the remaining demographic parameters fixed.

For each simulation, we classified genomic fragments according to their length and associated TMRCA values. We then summarized the resulting distributions in two-dimensional matrices, in which each cell represents the abundance of fragments within a specific combination of fragment length and TMRCA. To compare the introgressed and non-introgressed scenarios, we generated *cold plots*. These are

matrix-based visualizations, similar to heatmaps, that highlight the regions of the fragment-length and TMRCA space in which the two scenarios differ. In our analysis, the matrices were compared using two complementary approaches: multiplication and subtraction.

First, we multiplied the log-transformed matrices obtained under the introgressed and non-introgressed scenarios. This comparison revealed a region in the upper-right corner of the multiplication cold plots (Supplementary Figure 3.A), corresponding to fragments with long lengths and elevated TMRCA values. Under introgression scenario they become observable at deep divergence times, although this signal is weak and tends to disappear as the divergence time becomes more recent, becoming completely indecipherable at the suggested divergence point (~ 120 kya) between eOOA and main OOA human groups.

We then repeated the analysis by subtracting the matrix obtained under the non-introgressed scenario from the matrix obtained under the introgressed scenario. Positive values therefore indicate combinations of fragment length and TMRCA for which introgressed simulations contain more fragments than non-introgressed simulations. The subtraction cold plots showed an excess of fragments in the introgressed scenarios (Supplementary Figure 3.B). However, this excess was strongly dependent on divergence time: the signal decreased as the TMRCA between the source and recipient populations became more recent. These results indicate that, under the demographic conditions evaluated here, the ability to distinguish introgressed fragments is progressively reduced as the divergence time between the populations decreases.

**Analysis of real data**

Given these results, we sought to evaluate whether a basic eOOA scenario such as the considered could be detected in empirical data. We therefore modelled potential eOOA introgression into present-day Southeast Asian populations. The presence of early modern human-like fossils in East and Southeast Asia, together with the high levels of Denisovan admixture observed in these populations[5,29–31], led us to hypothesize that any detectable eOOA signal would most likely be found in

this region. To test this hypothesis, we conducted demographic modelling of eOOA introgression using the ABC-DL framework developed by Mondal et al.[21].

We used six present-day populations and two archaic hominin groups to model potential eOOA introgression. The present-day dataset included three Southeast Asian populations: the Temiar from Malaysia, the Kosipe from Papua New Guinea, and the Agta from the Philippines. These populations were selected as proxies for the genetic diversity of Southeast Asia. We also included Han Chinese individuals as a proxy for East Asian populations, allowing us to test whether eOOA introgression may have occurred before the divergence of East and Southeast Asian populations. In addition, CEU and Yoruba individuals were incorporated as proxies for European and sub-Saharan African populations, respectively. These populations enabled us to model the main Out-of-Africa event and to include an African outgroup without Neanderthal or Denisovan introgression. Finally, we incorporated Neanderthal and Denisovan genomes to account for other archaic introgression events and distinguish a potential eOOA contribution from ancestry derived from these hominin groups (Figure 5).

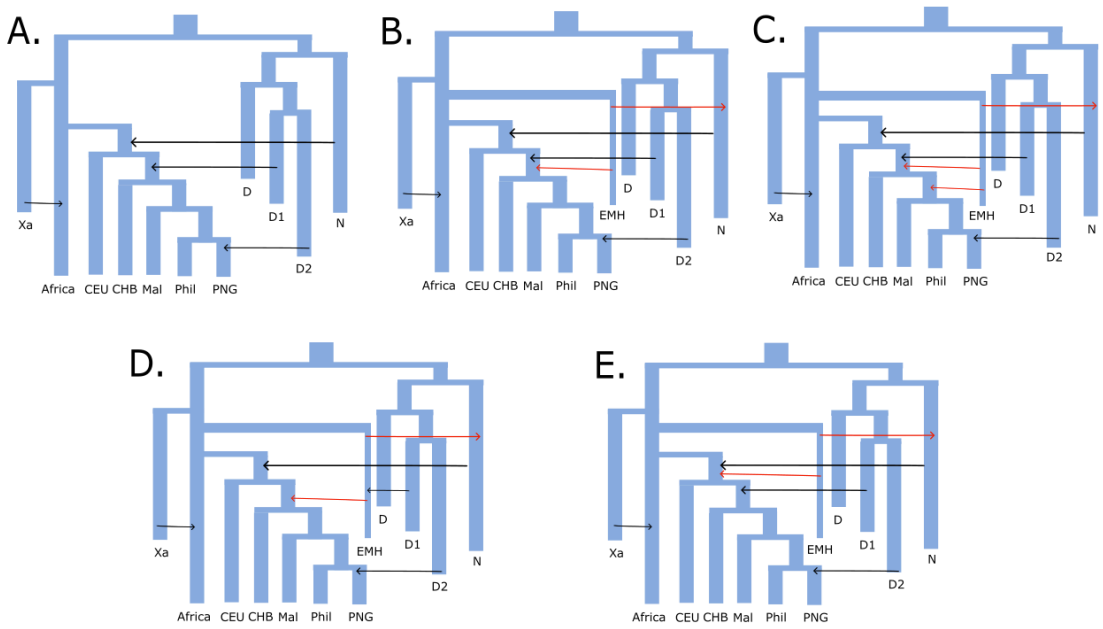



*Figure 5: Demographic models for ABC-DL analysis of introgression signal of early out-of-Africa into main out-of-Africa populations. Populations in models are: Xa (Ancestral ghost African), Africa (African Yoruba), CEU (Central European from*

*Utah), CHB (Chinese -Han from Beijing), Mal (Malaysian from Temiar), Phil (Philipines from Agta), PNG (Papuan New Guinea from Kosipe), EMH (Early Modern Human population – target population), D (Deniosva from Altai) and N (Neanderthal from Altai). Models B to E include early modern human admixture*.

For the population-structure analyses, we expanded the dataset by incorporating additional Southeast Asian populations and populations from the 1000 Genomes Project (see Materials and Methods; Supplementary Figure 4).

***Population structure***

To characterize the genetic structure of the dataset prior to demographic modelling, we performed principal component analysis (PCA) after standard quality-control filtering, yielding a final dataset of 163 individuals and approximately 10.5 million SNPs. Three PCA analyses were conducted using the complete dataset, a dataset excluding archaic genomes, and a dataset restricted to Southeast Asian populations. The first PCA recovered the expected population structure, with Neanderthals and Denisovans clearly separated from modern humans and African populations occupying the greatest genetic distance from the archaic genomes (Supplementary Figure 5A). Excluding archaic genomes highlighted the differentiation between African and non-African populations as well as the overall Southeast Asian cluster (Supplementary Figure 5B). Within Southeast Asia, Papuan populations formed a distinct cluster relative to the remaining Southeast Asian groups (Supplementary Figure 4), consistent with previous observations of their differentiated demographic history[32]. Overall, these analyses confirmed the expected population structure and supported the use of these populations for subsequent demographic modelling.

***Demographic modelling of eOOA introgression using whole genome data***

Five demographic models were evaluated to investigate potential locations of the eOOA introgression event into the main *Homo sapiens* lineage. Model A represents the null scenario, assuming no eOOA introgression, whereas Models B–E incorporate introgression events targeting different recipient ("sink") populations. Among these, Model D is the most distinct, as it uniquely includes an interaction

between Denisovan populations and the eOOA lineage (Figure 5). A detailed description of each model is provided in the Materials and Methods section.

We evaluate all competing models by means of the ABC-DL algorithm framework. In this algorithm we can distinguish competing demographic models using an optimal set of summary statistics extracted from the multi-Site Frequency Spectrum with the ABC-DL package[21]. For each model, we generated 200.000 simulations and analyzed the resulting summary statistic using the *abc* package in R[33].

This analysis complements the previous simulation-based assessment using fragment length and TMRCA distributions. While that approach relied on information closely related to the simulated introgressed fragments, the ABC analysis uses only the multi-SFS and therefore captures a more indirect signal of introgression. Consequently, part of the discriminatory power observed in the theoretical analysis is expected to be lost. Although statistics based on similar information—such as those derived from ARGs[34] could in principle provide substantially more genealogical information, their incorporation into an ABC framework would require reconstructing an ARG for every simulated dataset, making such an approach computationally impractical at the scale required for ABC.

First, we checked how well the ABC is able to identify between competing models by doing a cross-validation analysis of a subset of 250 models (50 each). The ABC struggles to differentiate between each of the competing models (Figure 5), as the highest cross validation is in Model D with a 0.74 identity probability, while Model A, the one without any introgression event, can be properly identified just slightly over 50% of the time (Table 1). The poor cross validation between introgressed models is expected, since the simulated models present few differences. More specifically, models A, B and E only differ in when a to which population the eOOA introgression occurs. After checking the ability of the method to differentiate between models, we checked which model is better supported by the data. For that, we obtained the best-fitting model using the *postpr* method in the *abc* package. The ABC supports

Model A (Figure 5.A) with around 69% of the accepted simulations identified as Model A, while Model D is only chosen 0.8% of the time. The other three models, which by the cross-validation can be treated as one, represent close to 30% of the accepted simulations (Table 2). Overall, with a Bayes factor of 2.4, Model A offers only anecdotal support over models considering eOOA introgression[35], precluding any definitive conclusions.

| | **Model A** | **Model B** | **Model C** | **Model D** | **Model E** |
|---|---|---|---|---|---|
| **Model A** | **0.5772** | 0.1236 | 0.1110 | 0.0516 | 0.1339 |
| **Model B** | 0.1077 | **0.3083** | 0.2592 | 0.0965 | 0.2284 |
| **Model C** | 0.0978 | 0.2740 | **0.2894** | 0.0624 | 0.2764 |
| **Model D** | 0.0279 | 0.0791 | 0.0758 | **0.7443** | 0.0728 |
| **Model E** | 0.1130 | 0.2337 | 0.2708 | 0.0808 | **0.3017** |

*Table 1: Cross-validation error of the ABC analysis with the cv4postpr method of the abc package. Rows correspond to the true model while columns correspond to the predicted model by the ABC. Model B, C and E show a very low self-identification, the abc is not able to differentiate between those three models.*

Together, the simulation and ABC analyses indicate that demographic models including early modern human introgression become difficult to distinguish from models without introgression when divergence between the interacting populations is shallow. Consequently, the lack of support for eOOA introgression in present-day genomes should not be interpreted as evidence that such admixture did not occur.

| **Model A** | **Model B** | **Model C** | **Model D** | **Model E** |
|---|---|---|---|---|
| **0.690** | 0.107 | 0.086 | 0.008 | 0.110 |

*Table 2: ABC discrimination process using the abc R package. Probability of selection of each model in the 1000 best simulations. Model A is the most selected model with 69% of 1000 best simulations.*

## Discussion

The apparent discrepancy between archaeological evidence for early dispersals of anatomically modern humans out of Africa and the demographic models inferred from present-day genetic diversity has been the subject of sustained debate within the scientific community[36–38]. The fossil record supports the occurrence of multiple early expansions of Homo sapiens beyond Africa, including sites in the Levant, Arabia, and East and Southeast Asia that predate the major 50–60 kya dispersal[11,14,16]. In contrast, genomic analyses consistently support a dominant later expansion as the principal source of ancestry in present-day non-African populations[1–3]. Population genomic models—largely based on genetic variation observed in living populations—have traditionally concluded that present-day human genetic diversity can be explained primarily by a single, later out-of-Africa expansion[39–42]. From this perspective, earlier dispersals, even if they did occur as indicated by archaeological evidence, are interpreted as having contributed little or nothing to the gene pools of modern populations[10,40,42,43].

Here, we tested whether the previous lack of evidence of a signal of an eOOA event in present-day human genomes could be due to not explicitly including the eOOA event in demographic models or due to the limited sensitivity and informativeness of summary statistics for detecting this introgression event under a human demographic scenario. To this end, we first evaluate the sensitivity of statistics based on fragment-length and TMRCA distributions for detecting eOOA introgression, as these quantities capture genealogical features that SMC-based methods are designed to model[26,44]. Specifically, we simulated a simplified human-like demographic scenario, varied the introgression parameters, and compared the resulting distributions of fragment lengths and their associated TMRCA values. And then, we used ABC-DL to evaluate alternative scenarios against real data involving eOOA introgression into populations located along the putative dispersal route, using sampled DNA sequences. This framework maximizes the informativeness of the site frequency spectrum (SFS) for discriminating among competing models[21,45].

## 1. Divergence time governs the detectability of introgression

In recent years, new methodological approaches focusing on inferring coalescent events have been proposed[21,26,34]. These methods do not rely exclusively on fitting predefined demographic models but instead examine patterns in genetic data more flexibly, sometimes focusing on signals of deep ancestry, unusual haplotype structure, or deviations from expected patterns under standard demographic scenarios, rapidly becoming the gold standard in population genomics. These approaches have been at the forefront of genomic inference, providing the resolution necessary to uncover ghost populations and disentangle admixture among closely related groups[46–48]. Such high-resolution frameworks are particularly essential given recent advances in reconstructing deep human history, which demonstrate that African and Eurasian lineages diverged earlier than 70 kya from a weakly structured, interconnected African metapopulation[49]. This complex demographic baseline is further extended by evidence that modern humans were shaped by repeated episodes of deep mixture—specifically extensive gene flow between expanding and local African groups around 50 kya, and the amalgamation of lineages diverged by 300,000 to 1 million years ago[48]. Consequently, isolating a genuine eOOA signal from this deep, admixed substructure demands exceptional analytical sensitivity[18].

We ran a set of simulations where we retrieved the TMRCA and fragment length of each simulated locus. A central outcome of our simulations is the explicit delineation of the demographic parameter space in which introgression from an early dispersal becomes detectable. In our framework, three parameters were varied: (i) divergence time between source and recipient populations (90 kya–2 My), (ii) admixture fraction (0–0.5 in exploratory simulations; 0–0.2 in realistic scenarios), and (iii) recombination rate ($2.5\times10^{-9}$–$2.5\times10^{-8}$ per base per generation). Detectable signals—defined as a consistent excess of long genomic fragments with elevated TMRCA relative to the non-introgressed expectation—emerged only when divergence times exceeded ~400–500 kya and admixture fractions were moderate to high (≥0.1–0.2). To quantify the separation between introgressed and non-introgressed fragment distributions, we calculated a weighted Jaccard distance

and modelled its relationship with TMRCA and introgression proportion using beta regression. Weighted Jaccard distance increased significantly with TMRCA, and the significant positive interaction between TMRCA and introgression proportion indicated that the effect of introgression became stronger as divergence time increased. At divergence times ≤200 kya, even admixture levels approaching 0.5 failed to produce a statistically separable signal. Within the more realistic prior space (divergence 90–700 kya; admixture 0–0.2), the detectable region was restricted to the upper boundary of divergence (>400 kya) combined with the highest admixture values. Recombination rate variation affected fragment counts but did not materially shift the detectability threshold. This mapping provides explicit boundary conditions for the genomic visibility of early dispersals.

Under plausible eOOA scenarios—divergence ≤200 kya and admixture likely ≤0.1–0.2—the parameter combinations fall outside the detectable region. Thus, early modern human dispersals temporally close to the main 60 kya expansion would require either substantially deeper prior isolation or implausibly high introgression proportions to leave a detectable genomic signature based on the TMRCA and fragment length statistics.

Overall, these results reflect a fundamental property of coalescent genealogies: when population splits are shallow relative to effective population size, most lineages coalesce before sufficient differentiation accumulates. Consequently, introgressed segments do not generate the elevated divergence tail characteristic of detectable archaic admixture. In contrast, Neanderthal and Denisovan introgression is identifiable precisely because those lineages diverged several hundred thousand years earlier[4,7]. Detectability based on coalescence times scales strongly—and nonlinearly—with divergence depth.

On top, it must be acknowledged that coalescent times can only be inferred (i.e. through statistics derived from ancestral recombination graphs (ARGs), which explicitly reconstruct local genealogical relationships across the genome[34]) or approximated by proxy summary statistics, such as Site Frequency Spectrum (SFS), in real data. While SFS have been a powerful statistic to reconstruct past population dynamics[50,51], it possesses fundamentally limited informativeness for resolving

443 complex or deeply divergent histories. This limitation arises not only because
444 distinct demographic scenarios can produce identical allele frequency spectra[52],
445 but also because the SFS inherently discards spatial relationships and linkage
446 disequilibrium (LD) between sites, thereby losing critical temporal resolution[53,54].
447 Because demographic estimates remain highly sensitive to these methodological
448 choices[55], capturing the true footprint of ancient introgression requires looking
449 beyond unlinked sites. However, incorporating ARG-based summary statistics into
450 an Approximate Bayesian Computation (ABC) framework would require
451 reconstructing an ARG for every simulated dataset, making such an approach
452 computationally prohibitive at the scale required for ABC analyses.

453 Therefore, under realistic divergence times and modest admixture fractions,
454 introgression from early dispersals would be genealogically indistinguishable from
455 background variation and thus leave no detectable signal in real data.
456 Consequently, genomic silence cannot be taken as evidence that early dispersals
457 did not occur; rather, it could reflect the theoretical limits of detectability under
458 shallow divergence. Our results demonstrate that, under realistic divergence times
459 and admixture fractions, introgression from closely related populations may be
460 mathematically indistinguishable from background variation. This has broader
461 implications for human evolutionary inference: genomic data seems to be
462 inherently biased toward detecting deeply divergent and/or high-admixture events,
463 such as Neanderthal and Denisovan introgression[4,7], while shallow and weakly
464 differentiated episodes may remain invisible. Consequently, the absence of
465 genomic evidence cannot be equated with demographic absence. Instead,
466 interpretations of human history must integrate theoretical detectability limits with
467 archaeological and paleoanthropological data, recognizing that some episodes of
468 expansion, interaction, or replacement may lie beyond the current resolution of
469 genomic inference.

470 2. ABC favors a model without detectable early Out-of-Africa introgression

471 Demographic models are typically designed to be parsimonious, aiming to explain
472 observed genetic patterns using the smallest possible number of demographic
473 parameters. While parsimony is a valuable principle for avoiding overfitting and

ensuring interpretability, it also limits the model's ability to detect demographic events that are not explicitly included in its structure[56]. If early dispersals out of Africa are not formally incorporated as candidate events, any genetic signatures resulting from them must be absorbed and explained by other parameters already present in the model, such as changes in effective population size or later migration events. Consequently, early dispersals may remain effectively invisible within the inferred demographic framework.

In our study, we considered a basic out of Africa demographic model and extended it by adding additional admixture events. Our ABC–DL framework consistently identified the model lacking any early Out-of-Africa (eOOA) introgression (Model A) as the best-fitting explanation of the observed genomic data. Posterior model probabilities indicate that Model A accounts for approximately 69% of the accepted simulations, whereas models incorporating eOOA gene flow receive substantially lower support, and Model D, which include intermediate admixture events, is selected in less than 1% of cases. Although cross-validation analyses reveal limited discriminatory power among closely related models—an expected outcome given their shared demographic structure—the overall posterior distribution remains decisively skewed toward the non-introgressed scenario.

These findings align with previous genomic studies supporting a dominant 50–60 kya dispersal as the primary source of ancestry in present-day non-African populations[1–3], and provide no statistical support for a substantial genetic contribution from earlier modern human expansions. Importantly, this result does not contradict the well-documented fossil evidence for early *H. sapiens* presence outside Africa[11,14,16], but rather indicates that, if such early populations contributed genetically to later groups, their contribution must have been either very low or genomically indistinguishable from background variation under current analytical frameworks.

Our simulation-based analysis provides an optimistic benchmark for the performance of genealogy-based methods because it relies on the true introgressed fragment boundaries and their corresponding coalescence times rather than inferred quantities. Despite having access to this perfect genealogical

information, the introgression signal remained undetectable across much of the parameter space explored. Consequently, although ARGs encode richer genealogical information than fragment length and TMRCA alone, approaches relying primarily on the inference of these same genealogical features are unlikely to substantially improve detection under these demographic scenarios. We therefore conclude that the lack of power reflects an intrinsic limitation of the available genealogical signal rather than the choice of summary statistic or inference framework which imply new methodological approaches to bypass this limitation.

### 3. Limited model discrimination reflects intrinsic demographic similarity rather than methodological failure

However, cross-validation analyses reveal limited power of our ABC-DL framework to distinguish among models that differ only in the presence or placement of an eOOA introgression event. This limitation stems from close demographic similarity between competing scenarios rather than from shortcomings of the ABC–DL framework. Because the putative eOOA and main OOA lineages diverge only shallowly on coalescent timescales, their expected summary statistics -based on SFS- overlap extensively, producing minimal contrast between models. When divergence time is recent relative to effective population size, introgressed alleles do not produce a distinct shift in the SFS but instead blend into existing variation. Under such conditions, even sophisticated frameworks such as ABC–DL, trained on large numbers of simulations, lack sufficient contrast to reliably separate competing demographic models. This pattern is consistent with previous simulation-based studies showing that detection of introgression from unsampled or “ghost” populations becomes increasingly difficult as divergence time decreases and admixture fractions remain modest[19,22,24]. Our findings therefore indicate that the limited model resolution reflects intrinsic genealogical proximity between scenarios rather than methodological failure.

### 4. Concluding perspective: Absence of signal does not imply absence of history

Taken together, our results indicate that demographic replacement does not necessarily imply genomic traceability. If early modern human populations outside Africa were later absorbed or largely replaced by the main 60 kya expansion, and if their divergence from the expanding lineage was shallow, their genetic contribution would blend into background variation. The detectability of introgression is bounded by fundamental genealogical constraints rather than by data quantity or methodological sophistication. The absence of detectable signal in present-day genomes, therefore, cannot be equated with the absence of earlier dispersal events.

This conclusion reframes the apparent tension between archaeology and genomics. Our findings show that fossil evidence of early dispersals and the lack of their fingerprint in genomic data are not mutually exclusive: demographic events can occur without leaving recoverable genomic traces. As a result, interpretations of human evolutionary history must incorporate the theoretical limits of detectability, recognizing that some episodes of expansion, interaction, and replacement may be permanently hidden within the genealogical noise of shared ancestry.

## Materials & Methods

### Defining the limits of introgression using simulated data

To delineate the limits of detectability of introgression, we generate simple demographic scenarios where we can observe how the introgression signal behaves at different divergence times, and different amounts of introgression.

### Simulating introgressed and non-introgressed dataset

We simulated the distribution of coalescent times of pairs of lineages of 1Mb size sequences under variations of a demographic model involving four populations: three representing the contemporaneous continental groups of human populations (African, European and Asian) and one more representing an archaic population that introgresses in the ancestor of these populations. We varied the parameters of time of the divergence of the archaic population and amount of introgression, also

considering different values of recombination, resulting in three thousand models. We fixed all parameters that are not directly influencing the gene flow process (effective population size of all populations and the split times of population 1, 2 and 3), we also fixed the moment of the introgression event to resemble a plausible eOOA introgression event where the eOOA population would introgress with an Asian group around after the split of Asian populations from the rest of Eurasia. Finally, we define prior distributions for the divergence time of the source population from the rest, the migration rate between source and sink populations and we also define different levels of recombination. Prior distributions of each of the tested parameters are the following; divergence time between source and rest: (90 kya – 2 My), migration rate between pop4 and pop3: (0-0.5), uniform recombination rate of the genome: (2.5e-9-2.5e-8).

We then compared the distribution of fragment size with TMRCA of the 3000 models with introgression with a null model with the same recombination rate, but without introgression. All simulations for the limits of introgression analysis have been conducted using fastSimcoal2[57] version 28, and the fixed parameters are all available in the supplementary data (Sup Table. 2).

**Preprocessing of simulation outputs**

The output we are interested in from fastSimcoal2 are the TMRCAs and the size of each of the fragments of a genome as the presence of big fragments with deep TMRCA could be an indicative introgression[58]. We have to take into account that after the introgression event, recombination could take place and break big fragments into smaller chunks that increase the difficulty in detecting introgressed fragments as they get masked by non-introgressed ones. To bypass this limitation, we group all adjacent fragments that present the exact same TMRCA, assuming that adjacent fragments with the same TMRCA come from the same introgression event, allowing us to increase the power to detect introgression in our simulations.

**Compare introgressed and non-introgressed simulations**

To identify how the introgression signal behaves depending on the amount of introgression and the TMRCA between the source and target populations, we

compare the distribution of genomic fragments of the target population under to demographic models with and without introgression across different TMRCAs and recombination rates. First, we observe how a change in the divergence time affect the amount of introgressed fragments we observe, so we fixed the amount of introgression to 0.5, meaning that 50% of the source ("ghost population") genome is present in the sink population at time 0, and modify our TMRCAs in between 90k-2My. The second scenario tested how the change in introgression rate affected our ability to observe, for this we fixed all simulation to an introgression rate of 2My and allow simulation to change between 0.01 to 0.5 introgression rate. In both scenarios we replicated the analysis with different levels of uniform recombination, as the amount of recombination could mask our ability to detect bigger fragments that could be identified as introgressed.

Correlation of weighted Jaccard distance between introgressed and non-introgressed fragment distributions. The Jaccard distance is a measure of dissimilarity between two datasets. It is derived from the Jaccard similarity index[59], which is calculated as the ratio of the intersection to the union and ranges from 0 to 1. In our analysis, we adapted this metric to obtain a weighted Jaccard distance that quantifies the dissimilarity between the distributions of fragments assigned to the ancestral population (*deme* = 0) under introgressed and non-introgressed models. The weighted version assigns greater importance to regions of the distribution containing a higher proportion of fragments. We then calculated the Spearman correlation between the weighted Jaccard distance and either TMRCA or the introgression rate using one hundred simulations in which one parameter was fixed while the other varied. After identifying correlations with both TMRCA and the introgression rate, we used beta regression to model their individual effects, as well as their interaction, on the weighted Jaccard distance.

**Detect introgression under more realistic scenarios**

More simulations were produced with fastSimcoal2[57] v28, where the interested parameters fluctuated between more realistic scenarios. All parameters of interest are not fixed but rather generate from a prior distribution for each parameter that has a direct influence on our ability to detect introgression. Therefore, having all the

other demographic variables fixed we define prior distributions for the amount of introgression (0 - 0.20), the TMRCA between the source and target populations (90Ky - 700Ky) and the recombination rate (2.5e-9 - 2.5e-8). From this set of simulations, we repeat the observation of the distribution of genomic fragment length compared to the TMRCA of those fragments and we perform a cold spot analysis to observe regions in the space that could be relatively increased under an introgression scenario.

For the cold plot analysis, we first transform the introgressed and non-introgressed matrices obtained from the TMRCA-fragment length distribution plots, by scaling them by the mean number of fragments in each bin in the matrix, defining 0.2 size bins. Once we have both matrices scaled, we multiply the introgressed matrices by the non-introgressed one, the result is going to be a matrix where negative values represent regions of bin size 0.2 where one of the scenarios has an increased representation of fragments compared to the mean number of fragments observed in each bin. We then can observe which matrix has a positive value and with that we can observe if there is an increase of fragments in the introgressed scenario with respect to what is expected under a non-introgressed one. We also subtract the non-introgressed matrix from the introgressed one to observe if there are bins with more introgressed fragments in absolute values compared to the no-introgression model.

Assessing Early Modern Human introgression from whole genome data

To analyze the possible effect of an Early Modern Human introgression into current day individuals, we applied an Approximate Bayesian Computation coupled with Deep Learning (ABC-DL)[21] to test whether a demographic model with an introgression event from an early out-of-Africa (eOOA) population better represents current human datasets than a model that does not include an introgression event from that eOOA group.

Dataset

First, we define the individuals that are going to be used as observed data for the ABC-DL analysis. We collected a dataset of 172 individuals from East, South East

Asian and Oceanian groups plus references of the surrounding major groups (Sup Table 1). Our dataset includes 90 whole genome samples from 10 different South East Asia and Oceanian populations (Philippines (Aeta = 16, Batak = 12, Agta = 11, Mamanwa = 6) and Malaysia (Jehai = 9, Kintaq = 6, Temiar = 6) as well as different Indigenous groups from Papua New Guinea (Koinambe = 12, Kosipe = 10, Australia = 2)) sequenced with Illumina HiSeq X technology, obtained from libraries constructed with the 10X Genomics Chromium technology. These samples are available upon request at the European Genome-phenome Archive (EGA) with study accession number EGAS50000000044. Four current day reference individuals (10X Papuan from Highlands = 2 and Sepik = 2) come from Bergström et al.[60], 17 South and South East Asians (4 Jarawa, 6 Onge, 2 Birhor, 3 Irula, 2 Riang) from Mondal et al.[20], 4 Malaysian (Jehai) from Aghakhanian et al.[32], 13 from 1000 Genomes Project[61] and 36 individuals from Simons Genome Diversity Project[62] while the ancient reference genomes are from Moreno-Mayer et al. (USR1)[63], Sikora et al. (AHUR = 1, Yana = 1, Sumidouro 5 = 1)[43], Fu et al. (Ust-Ishim)[1] as well as Neanderthal (Altai Neanderthal) and Denisova (Altai Denisova) samples from Püfer et al.[7] and Meyer et al.[64] respectively. Fully collected dataset is available upon request as a VCF file mapped to the hg37 version of the human genome.

Data Preprocessing

We obtained the structure dataset by merging all samples, keeping the common SNPs across all samples, and applying SNP (geno = 0.1) and individual missingness (mind = 0.1) in PLINK v2[65]. We then excluded the SNPs with MAF < 0.01 and removed the individuals with cryptic relatedness using the KING (KING) method from PLINK v2 (PLINK) based on a cutoff of 0.325. The final dataset comprised ~10.5 M SNPs on 172 individuals from 56 different populations.

For the demographic analysis we further process the structure dataset by including the ancestral state of each variant by referencing it to the chimpanzee genome panTro4 genome assembly from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/vsPanTro4/reciprocalBest to avoid any discrepancy between African and non-African populations as detailed in Mondal et al.[21].

Structure analysis

**Principal component analysis**

Principal component analysis (PCA) was performed using SNPRelate[66] in R[67], pruning the data for linkage disequilibrium between the markers using PLINK v2.0[65] based on an r2 threshold of 0.4 in every continuous window of 200 SNPs with a step of 25 SNPs (i.e., –indep-pairwise 200 25 0.4). Three PCAs are conducted with different sets of data, the original dataset (163 individuals), one dataset without non-Homo sapiens individuals (160 individuals), and another dataset including only samples from South East Asia & Oceania (119 individuals).

Demographic modelling of eOOA introgression

To investigate the potential presence and effect of eOOA introgression in present-day populations, we employed an Approximate Bayesian Computation–Deep Learning (ABC-DL) approach as described in detail by Mondal et al.[21]. The ABC-DL implementation consisted of three main steps. First, we generated thousands of simulations for each competing demographic model using fastsimcoal2[57] (v26X), with the joint multidimensional site frequency spectrum (jSFS) across populations serving as the raw summary statistic. The jSFS captures information relevant to most frequency-based population genetic analyses and provides insights into the demographic parameters modeled (see Mondal et al., 2019[21]). Second, we trained a deep learning (DL) model to predict, from the jSFS, the most informative summary statistic (SS-DL) for each parameter or model comparison. A potential limitation of this approach lies in the risk of overfitting, as the DL model is trained on simulated data and subsequently applied to empirical data. To minimize bias in DL predictions, we introduced jSFS noise from the observed data into each simulation, following the procedure described by Mondal et al. 2019[21]. Finally, a classical ABC analysis was performed using the DL-derived summary statistic (SS-DL) on a new set of simulated datasets.

The callable fraction of the genome used for demographic modeling was based on a modified version of that defined by Pouyet et al.[68]. To identify neutral regions, we further refined the callable genome by masking all Ensembl-annotated genes within

a 20-kb flanking range and CpG islands as defined in Mondal et al.[21]. Regions failing to reach a SNP density above 90% in 10-kb windows (sliding step of 2.5 kb) were subsequently excluded, yielding a total of 5475 Mb of callable genome used for analysis.

Ten deep learning networks were constructed, each with four hidden layers, and trained with 20,000 simulations per network. An additional 180,000 simulations per model were generated by injecting jSFS noise derived from observed data, using different combinations of individuals from the populations included in the demographic models (4 Europeans, 4 CHB, 4 YRI, 21 SEA, one Neanderthal, and one Denisovan). Model probabilities were predicted using the ten DL networks, and results were averaged to obtain the final SS-DL for use in the ABC analysis. The ABC computations were conducted in R with the abc package[33,67], assuming a mutation rate of $1.61\times10^{-8}\pm0.13\times10^{-8}$ per site per generation[69] and a generation time of 29 years[70].

The ABC analysis was carried out in two steps. First, we evaluated how effectively the method could distinguish among competing models using the cv4postpr function in the abc package, which performs cross-validation by treating simulated datasets as observed data. For each model, 50 simulations were used with the same ABC parameters as those applied to the observed data, and the number of times the model with the highest posterior probability corresponded to the true generating model was recorded. Once discrimination between models was confirmed, we proceeded to select the best-fitting model using the postpr function of the abc package. The 1,000 best simulations (out of 180,000 per model) were retained under the "mnlogistic" option.

**Model Description**

We define 5 different models with different potential locations of the introgression event of eOOA towards the "main" *Homo sapiens* lineages. The first models, known as Model A, includes the South East Asian and Oceanian population as a sister clade of East Asia, also European and African populations as the "main" *H. sapiens* line. It also includes a Neanderthal population that introgressed towards the

common ancestor of European, East Asian, South East Asian and Oceanian and Denisova populations (3 different Denisova groups as per Ongaro and Huerta-Sanchez[71], that introgressed in two admixture pulses towards East Asia and Papuan respectively. In Model A we don't include any eOOA introgression event as it would serve us as a basal model to test against the introgressed models (Figure 5.A). Next, we have Model B (Figure 5.B), this model includes all the same parameters as Model A but adding an eOOA (defined as a ghost population) event around 130k years ago that introgressed twice into surrounding populations. First, an early introgression event to Neanderthal followed by a later introgression into the common ancestor of all Asian populations. Model C (Figure 5.C) follows the same structure as Model B but includes an extra introgression event in South East Asia populations, increasing the amount of ancestry present in South East Asia from this eOOA group. Model D (Figure 5.D), poses a different variation of the scenario in Model B. In this model the introgression event from Denisova does not go directly to the "main" clade but introgresses with the eOOA population that later introgressed to the ancestors of the Asian clades. Finally, Model E (Figure 5.E) presents the eOOA introgression deeper in the *H. sapiens* differentiation process. In this model, the introgression event comes before the split between European and Asian populations, in an event that could be related to some basal Eurasian admixture process.

The specific parameters in the prior distributions have been defined following information obtained from references[2,30,42]. The specific prior distributions for each parameter are available in the supplementary table (S. Table.2).

**Acknowledgements**

We thank Moises Coll and Manuel Ferrando for helpful comments, discussions and ideas. We would also like to thank Marta M Lahr for posing anthropological questions and genetic problems not always easy to solve; maybe this paper will not meet her expectations.

**Author contributions**

J.S.: Conceptualization, data analysis, investigation, methodology, visualization, writing original draft. A.N.: Conceptualization, validation, data analysis. J.B.:

Conceptualization, investigation, writing the original draft. O.L.: Conceptualization, validation, methodology, investigation, writing original draft. All authors contributed to the manuscript revision.

**Funding**

JM is hired under the Generation D initiative, promoted by Red.es, an organization attached to the Ministry for Digital Transformation and the Civil Service, for the attraction and retention of talent through grants and training contracts, financed by the Recovery, Transformation and Resilience Plan through the European Union's Next Generation funds. Funding for JB: MICIU/AEI/10.1309/501100011033. And the "Unidad de Excelencia Maria de Maeztu" CEX2024-001431-M, funded by MICIU/AEI/10.13039/501100011033.

**Competing Interest**

Authors declare no competing interest.

**Data Availability**

Genomic data from SE populations are available under controlled access due to restrictions imposed by ethical approval and the informed consent from participants at the EGA study accession number EGAS50000000044. VCFs used in analysis available upon request.

**Code Availability**

Code is available at Github: albanieto/missigno & josm18/eOOA

**References**

1. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).

2. Malaspinas, A.-S. *et al.* A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016).

3. Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L. & Skoglund, P. Origins of modern human ancestry. *Nature* **590**, 229–237 (2021).

4. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).

5. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).

6. Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The Date of Interbreeding between Neandertals and Modern Humans. *PLOS Genetics* **8**, e1002947 (2012).

7. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

8. Stringer, C. B., Grün, R., Schwarcz, H. P. & Goldberg, P. ESR dates for the hominid burial site of Es Skhul in Israel. *Nature* **338**, 756–758 (1989).

9. Grün, R. *et al.* U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *Journal of Human Evolution* **49**, 316–334 (2005).

10. Bae, C. J., Douka, K. & Petraglia, M. D. On the origin of modern humans: Asian perspectives. *Science* **358**, eaai9067 (2017).

11. Hershkovitz, I. *et al.* The earliest modern humans outside Africa. *Science* **359**, 456–459 (2018).

12. Martinón-Torres, M., Wu, X., Bermúdez de Castro, J. M., Xing, S. & Liu, W. Homo sapiens in the Eastern Asian Late Pleistocene. *Current Anthropology* **58**, S434–S448 (2017).

13. Westaway, K. E. *et al.* An early modern human presence in Sumatra 73,000–63,000 years ago. *Nature* **548**, 322–325 (2017).

14. Liu, Y., Mao, X., Krause, J. & Fu, Q. Insights into human history from the first decade of ancient human genomics. *Science* **373**, 1479–1484 (2021).

15. Groucutt, H. S. *et al.* Rethinking the dispersal of *Homo sapiens* out of Africa. *Evolutionary Anthropology* **24**, 149–164 (2015).

16. Harvati, K. *et al.* Apidima Cave fossils provide earliest evidence of Homo sapiens in Eurasia. *Nature* **571**, 500–504 (2019).

17. Sümer, A. P. *et al.* Earliest modern human genomes constrain timing of Neanderthal admixture. *Nature* **638**, 711–717 (2025).

18. Tournebize, R. & Chikhi, L. Ignoring population structure in hominin evolutionary models can lead to the inference of spurious admixture events. *Nature Ecology & Evolution* **9**, 225–236 (2025).

19. Vernot, B. & Akey, J. M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017–1021 (2014).

20. Mondal, M. *et al.* Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat Genet* **48**, 1066–1070 (2016).

21. Mondal, M., Bertranpetit, J. & Lao, O. Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nat Commun* **10**, 246 (2019).

22. Villanea, F. A. & Schraiber, J. G. Multiple episodes of interbreeding between Neanderthal and modern humans. *Nat Ecol Evol* **3**, 39–44 (2019).

23. Durvasula, A. & Sankararaman, S. Recovering signals of ghost archaic introgression in African populations. *Science Advances* **6**, eaax5097 (2020).

24. Skov, L. *et al.* The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* **582**, 78–83 (2020).

25. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* **108**, 11983–11988 (2011).

26. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

27. Liang, M. & Nielsen, R. The Lengths of Admixture Tracts. *Genetics* **197**, 953–967 (2014).

28. Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sánchez, E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet* **16**, 359–371 (2015).

29. Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S. & Akey, J. M. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* **173**, 53-61.e9 (2018).

30. Jacobs, G. S. *et al.* Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell* **177**, 1010-1021.e32 (2019).

31. Teixeira, J. C. & Cooper, A. Using hominin introgression to trace modern human dispersals. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15327–15332 (2019).

32. Aghakhanian, F. *et al.* Unravelling the Genetic History of Negritos and Indigenous Populations of Southeast Asia. *Genome Biology and Evolution* **7**, 1206–1215 (2015).

33. Csilléry, K., François, O. & Blum, M. G. B. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol* **3**, 475–479 (2012).

34. Speidel, L. *et al.* Inferring Population Histories for Ancient Genomes Using Genome-Wide Genealogies. *Mol Biol Evol* **38**, 3497–3511 (2021).

35. Deines, Z. Using Bayes to get the most out of non-significant results. *Frontiers in Psychology* **5**, (2014).

36. López, S., van Dorp, L. & Hellenthal, G. Human Dispersal Out of Africa: A Lasting Debate. *Evol Bioinform Online* **11**, 57–68 (2016).

37. Montinaro, F., Pankratov, V., Yelmen, B., Pagani, L. & Mondal, M. Revisiting the out of Africa event with a deep-learning approach. *Am J Hum Genet* **108**, 2037–2051 (2021).

38. Rose, J. *et al*. A failed human expansion out of Africa 100,000 years ago. Preprint at https://doi.org/10.21203/rs.3.rs-8751643/v1 (2026).

39. Relethford, J. H. Models, predictions, and the fossil record of modern human origins. *Evolutionary Anthropology: Issues, News, and Reviews* **8**, 7–10 (1999).

40. Stoneking, M. & Krause, J. Learning about human population history from ancient and modern genomes. *Nat Rev Genet* **12**, 603–614 (2011).

41. Reyes-Centeno, H. Out of Africa and into Asia: Fossil and genetic evidence on modern human origins and dispersals. *Quaternary International* **416**, 249–262 (2016).

42. Mondal, M. *et al.* Resolving out of Africa event for Papua New Guinean population using neural network. *Nat Commun* **16**, 6345 (2025).

43. Sikora, M. A Genomic View of the Pleistocene Population History of Asia. *Current Anthropology* **58**, S397–S405 (2017).

44. Peede, D. *et al.* Not Just Ne Ne-More: New Applications for SMC from Ecology to Phylogenies. *Genome Biology and Evolution* **1**, (2025).

45. Serradell, J. M., Lorenzo-Salazar, J. M., Flores, C., Lao, O. & Comas, D. Modelling the demographic history of human North African genomes points to a recent soft split divergence between populations. *Genome Biol* **25**, 201 (2024).

46. Cousins, T., Scally, A. & Durbin, R. A structured coalescent model reveals deep ancestral structure shared by all modern humans. *Nat Genet* **57**, 856–864 (2025).

47. Speidel, L. *et al.* High-resolution genomic history of early medieval Europe. *Nature* **637**, 118–126 (2025).

48. Loya, H., Hinch, A. G., Palamara, P. F., Speidel, L. & Myers, R. Genome-wide genealogies reveal deep admixtures forming modern humans. Preprint at https://doi.org/10.64898/2026.04.17.719197 (2026).

49. Ragsdale, A. P. *et al.* A weakly structured stem for human origins in Africa. *Nature* **617**, 755–763 (2023).

50. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genet* **5**, e1000695 (2009).

51. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**, e1003905 (2013).

52. Myers, S., Fefferman, C. & Patterson, N. Can one learn history from the allelic spectrum? *Theoretical Population Biology* **73**, 342–348 (2008).

53. Terhorst, J. & Song, Y. S. Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7677–7682 (2015).

54. Omarjee, A., Lambert, A. & Achaz, G. Blockbuster meets the theoretical limit of genome-wide SFS-based inference of recent demography. Preprint at https://doi.org/https://doi.org/10.64898/2026.03.16.712122 (2026).

55. Beichman, A. C., Phung, T. N. & Lohmueller, K. E. Comparison of Single Genome and Allele Frequency Data Reveals Discordant Demographic Histories. *G3 (Bethesda)* **7**, 3605–3620 (2017).

56. Maier, R. *et al.* On the limits of fitting complex models of population history to f-statistics. *eLife* **12**, e85492 (2023).

57. Excoffier, L. *et al. fastsimcoal2* : demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021).

58. Huang, X., Hackl, J. & Kuhlwilm, M. Decoding genomic landscapes of introgression. *Trends in Genetics* **41**, 1096–1108 (2025).

59. Kosub, S. A note on the triangle inequality for the Jaccard distance. *Pattern Recognition Letters* **120**, 36–38 (2019).

60. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).

61. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

62. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).

63. Moreno-Mayar, J. V. *et al.* Early human dispersals within the Americas. *Science* **362**, eaav2621 (2018).

64. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**, 222–226 (2012).

65. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015-0047–8 (2015).

66. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).

67. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2023).

68. Pouyet, F., Aeschbacher, S., Thiéry, A. & Excoffier, L. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *Elife* **7**, e36317 (2018).

69. Lipson, M. *et al.* Calibrating the Human Mutation Rate via Ancestral Recombination Density in Diploid Genomes. *PLoS Genet* **11**, e1005550 (2015).

70. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).

71. Ongaro, L. & Huerta-Sanchez, E. A history of multiple Denisovan introgression events in modern humans. *Nat Genet* **56**, 2612–2622 (2024).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialeOOA.pdf