Article

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

Joaquim Fort [1,2] & Joaquim Pérez-Losada[1]

The Neolithic (i.e., farming and stockbreeding) spread from the Near East across Europe since about 9000 years before the common era (BCE) until about 4000 yr BCE. It followed two main routes, namely a sea route along the northern Mediterranean coast and an inland one across the Balkans and central Europe. It is known that the dispersive behavior of farmers depended on geography, with longer movements along the Mediterranean coast than along the inland route. In sharp contrast, here we show that for both routes the percentage of farmers who interbred with hunter-gatherers and/or acculturated one of them was strikingly the same (about 3.6%). Therefore, whereas the dispersive behavior depended on the proximity to the Mediterranean sea, the interaction behavior (incorporation of hunter-gatherers) did not depend on geographical constraints but only on the transition in the subsistence economy (from hunting and gathering to farming) and its associated way of life. These conclusions are reached by analyzing the clines of haplogroup K, which was virtually absent in hunter-gatherers and the most frequent mitochondrial haplogroup in early farmers. Similarly, the most frequent Y-chromosome Neolithic haplogroup (G2a) displays an inland cline that agrees with the percentage of interbreeding reported above.

The Neolithic transition is defined as the change from a hunting-gathering economy to one based on farming and/or animal husbandry. This major shift in human prehistory led to increasing food storage and division of labor. It also sparked the evolution towards new forms of human behavior in many other realms, such as urbanization, writing, exact numeracy, and mathematics[1,2].

The Neolithic originated in the Near East and spread from there across Europe[3,4] following an inland route (over the Balkans and central Europe) and a sea route (along the northern Mediterranean coast), as illustrated in Fig. 1a[5,6]. Two possible models (or a combination of them) have been proposed to explain these spreads. The demic diffusion model assumes that populations of farmers expanded their range,

cultivating new lands and therefore bringing agriculture with them[7]. The cultural diffusion model, on the contrary, proposes that there were no substantial population movements, and autochthonous hunter-gatherers (HGs) became farmers due to interbreeding between both populations and/or acculturation of HGs[8]. During the last decade of the previous century, some authors analyzed modern DNA data and found them consistent with a mainly demic diffusion spread of the Neolithic using a variety of approaches, such as calculating genetic distances and comparing them to spatial simulations[9], estimating spatial correlograms and divergence times[10], applying admixture models and spatial analyses[11] or calculating the proportion of Neolithic lineages using spatial simulations[12]. More recently, the technical

[1]Complex Systems Laboratory, Universitat de Girona, C/ Maria Aurèlia Capmany 61, 17003 Girona, Catalonia, Spain. [2]Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 3, 08010 Barcelona, Catalonia, Spain. e-mail: joaquim.fort@udg.edu
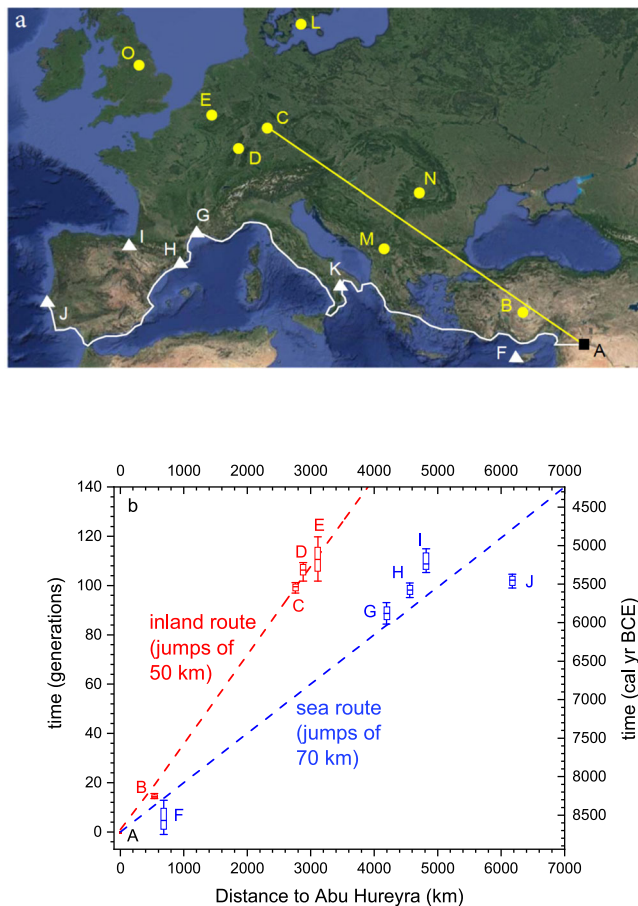
Fig. 1 | The two routes of Neolithic spread in Europe. a Visualization of the two routes. Each yellow circle is the oldest Neolithic site in a region reached through the inland route. Each white triangle is the oldest Neolithic site in a region reached mainly along the sea route. Distances are estimated using great circles for the inland route (e.g., yellow line) and along the coast for the sea route (e.g., white line). The black square A is Abu Hureyra (Syria), the presumed origin of the Neolithic wave of advance (see Methods, Initial conditions). **b** Arrival times of Neolithic farmers in several regions according to the radiocarbon dates of sites in **a** (Supplementary Data 4) and simulations (lines) for the inland route (red, jumps of 50 km per generation) and the sea route (blue, jumps of 70 km per generation). Each symbol is the radiocarbon date of the oldest Neolithic site in the region considered. The regions are A northern Mesopotamia, B Anatolia (present-day Turkey), C Germany, D northern France, E Belgium, F Cyprus, G southern France, H Catalonia, I Navarre and J central Portugal. Some regions (K Italy, L Sweden, M Serbia, N Romania, and O United Kingdom) are not included in **b** due to the existence of delays in the arrival of the Neolithic[5], but more complicated models accounting for them would yield the same conclusions (see Supplementary Information, Sec. S3-C). We have checked that the lines in **b** do not change appreciably for values of the interbreeding intensity η such that $\eta < 0.1$. The boxes (10% and 90% quartiles), error bars or whiskers (1% and 99% percentiles), and horizontal lines (medians) have been obtained using OxCal 4.4 with the calibration curve IntCal 20 (Supplementary Data 4). Source data are provided in the excel Source Data file.

possibility to analyze ancient DNA unveiled large genetic differences between local HGs and early Neolithic farmers in central Europe and this confirmed that demic diffusion played a primary role in the spread of the Neolithic[13]. The primacy of demic over cultural diffusion has been furthermore attested by genome-wide studies[14,15], which have shown that >90% of the genetic ancestry of early European farmers is due to a main population source in Anatolia (present-day Turkey)[14]. However, these studies do not quantify some key parameters describing the behavior of ancient individuals, including the following. What was the average distance moved per generation by pioneering farmers? Was it different along the central European and

Mediterranean routes? What was the proportion of early farmers that interbred with HGs? Was it different along both routes? Such questions need other approaches than the usual genetic and genomic analyses cited above.

The average distance moved per generation along each Neolithic route has been previously estimated by comparing archeological data to spatial simulations of Neolithic spread[4,16,17]. In contrast, the interaction behavior (interbreeding and/or acculturation) cannot be inferred using only archeological data because an extremely precise range for the spread rate would be necessary (Fig. 3a in ref. 18). For this, genetic data are also needed. Mitochondrial (mt) DNA is located outside the cell nucleus and passed (with mutations) from each mother to her children. Each individual has a mt haplotype, defined as a list of mt nucleotides. A mt genetic marker (also called a haplogroup) can be defined as a set of haplotypes that are thought to have a shared ancestor not shared with other mt haplotypes. During the last five decades, some authors have used numerical simulations to study genetic clines (i.e., the frequencies of genetic markers as a function of position) established during the Neolithic spread[9,11–13,19] (for a review of previous simulations, see Supplementary Information, Sec. S1-A). However, when those studies were published, none or a few ancient genetic data were available, so it was not possible to compare the simulated clines to observed ones. Such a comparison was finally performed seven years ago[20]. Here we take advantage of the substantial increase in the number of data available during the last seven years, which is now enough to estimate mt haplogroup frequencies using data from 961 early farmers who lived in 16 different geographical regions (in comparison, only 249 early farmers from 9 regions could be analyzed in ref. 20). This increase in the data available makes it now possible to analyze the interaction behavior between early incoming farmers and local HGs along the central European and Mediterranean routes separately (Methods, Observed data). Before doing so, we explain in more detail the rationale of our approach. We shall refer to interbreeding for clarity, but all remarks on interbreeding are also valid for acculturation or a combination of both because essentially the same mathematical equations can be used to describe both processes (Methods).

Some haplogroup-based analyses of modern populations led to misleading inferences on prehistory[21]. For example, Basque populations were proposed to have been isolated since the Paleolithic on the basis of modern haplogroup data[22], but ancient DNA later showed that their isolation began only in the Iron Age[23]. This problem does not affect the work reported here, because we use only ancient genetic data. Haplogroup studies using ancient DNA have yielded very important and sound observations, such as a clear genetic discontinuity between local HGs and the first farmers in central Europe[24], a resurgence of HG lineages during the middle Neolithic in the same region[25], etc. Moreover, the usefulness of single genetic markers has been shown by several recent ancient DNA studies, for example, analyzing mt haplogroup K[20] and Y-chromosome haplogroup H2[26] in the Early Neolithic, mt haplogroup H in the Middle and Late Neolithic[27], mt haplogroups A2 and C1 in pre-contact Caribbean islanders[28], etc. In this work, we are especially interested in two issues that call for the study of single genetic markers. First, it is appealing to find a quantitative explanation for the observed frequency cline of a genetic marker. Second, it is possible to apply the analysis of a single marker to estimate a basic interbreeding parameter of farmers and HGs, namely the percentage of farmers that interbred with HGs[20]. For both purposes, the usefulness of analyzing a single genetic marker can be understood by considering a genetic marker, the frequency of which is not affected by either drift or selection during the Neolithic spread. If local HGs lack this marker but some farmers have it, and there is admixture between both populations, the marker will dilute progressively as the population front of farmers propagates and autochthonous HGs admix with them, so the marker frequency will decrease (non-linearly[11]) with

increasing distance from the origin of the dispersal (see, e.g., pp. 82–83 in ref. [29]). Obviously the larger the percentage of farmers involved in interbreeding, the stronger the dilution of the Neolithic marker in farmers, i.e., the cline of this single marker will be steeper. Thus, such a single-marker cline can yield information on the interactive behavior of farmers and HGs. The existence of this interaction is not in question, because estimations of genomic ancestry and admixture timing have proven conclusively the existence of interbreeding between farmers and HGs since the arrival of the first farmers or shortly thereafter[15,30–34].

From the discussion in the previous paragraph, it follows that a specific genetic marker that meets the following conditions would be ideal for a quantitative study aiming to estimate the percentage of farmers that interbred with HGs: (1) the marker is (nearly) absent in HGs before the arrival of the firsts farmers; (2) selection and (3) drift effects (including surfing) are not important; and (4) the regional frequency of the marker considered reaches sufficiently high values so that a clear gradual variation can be detected (if it exists). The reason of condition (4) is that this detection is not possible for low-frequency markers, due to the small regional numbers of farmers whose haplogroup is known at present (see Supplementary Information, Sec. S11). If these four conditions are met, it is possible to compare the observed cline to those simulated by simple spatial demic-cultural diffusion models (as done below). Haplogroup K is the only mt haplogroup that satisfies all four conditions stated above (as shown in detail in Supplementary Information, Sec. S1-C). Concerning condition (1), although it is known that a few HGs who lived in Europe before the arrival of farmers have subclades of haplogroup K that have also been found in early farmers, taking them into account in our simulations does not change appreciably the results presented below, so our conclusions are maintained (Supplementary Fig. 1a).

The similarities and differences between the inland and Mediterranean routes of Neolithic spread in Europe are topics of interest, both archeologically[5] and genetically[35]. A well-known cultural difference is pottery, with Linearbandkeramik (LBK) ceramics present only in part of the inland route and Impressa/Cardial ceramics found only in some regions along the Mediterranean route. Another difference is that radiocarbon dates combined with spatial simulations have shown that the distances moved by early farmers each generation were longer along the Mediterranean than along the inland route[6,16,17]. Two more differences are a drop in early Neolithic crop diversity and a slowdown in the Neolithic spread rate along the inland route but not along the Mediterranean one[36]. On the other hand, early farmers from both routes have similar genomes, as expected if they descend from a single population that diverges along both routes[35]. However, there are clear genomic differences between both routes. Indeed, the proportion of HG ancestry in early farming populations is lower along all of the inland route and in the eastern part of the Mediterranean route, compared to the western part of the latter (which has a regional maximum in southern France)[31,32,37]. These differences between both routes are also reflected in principal-component maps, which make it possible to visualize the degree of similarity between populations of early farmers and HGs[31,38,39]. Very recently, another interesting analysis has focused on a specific genetic marker (Y-chromosome H2) in early European farmers and identified two subclades, called H2d and H2m, which are found respectively along the inland and Mediterranean routes of Neolithic spread[26]. The importance of comparing the farmer-HG interactions along both routes has been recently stressed[32,37,40], but the proportion of farmers that interbred with HGs and/or acculturated one of them along each route has never been estimated. The aim of the present paper is precisely to compare this feature of ancient behavior along both routes. Our approach is based on computing the frequency of mt haplogroup K in populations early farmers and comparing its dependence on distance to that predicted by a sound mathematical theory[20,41]. This makes it possible to estimate a key parameter in human prehistoric behavior, namely the percentage of early farmers that interbred with HGs and/or acculturated a HG (Methods).

Results

Before analyzing what the ancient cline of haplogroup K tells us about the interactions between the incoming farmers and the autochthonous HGs, it is necessary to develop a simulation model that provides reasonable overall agreement with the archeological data. In order to do so, in Fig. 1b we compare the date of the oldest Neolithic site in several regions (error bars, from Supplementary Data 4) to the arrival time of the Neolithic at the same site according to our simulations (lines). The latter times have been obtained by simulating the spread of farmers on a grid of square cells dispersing from an initial cell, reproducing and interacting with HGs which initially live on the rest of cells (Methods and Supplementary Information, Secs. S2 and S3). In order to find realistic results, we have considered two simulation models. First, the inland model is defined as that in which the fraction of individuals that moves has displacements of 50 km per generation on average, as suggested by ethnographic data[42] and used in previous simulations[16,20,41,43]. Using this inland model, we have obtained the red line in Fig. 1b. It gives the arrival time of the Neolithic as a function of the great-circle distance from the oldest site, which is located in northern Mesopotamia (map in Fig. 1a). We assume that this site (Abu Hureyra) is a reasonable estimation of the origin of the Neolithic wave of advance (see Methods, Initial conditions). We note in Fig. 1b that the inland model (red line) yields results that widely agree with the archeological data for regions shown as red error bars (Anatolia, Germany, northern France, and Belgium). However, this inland model (red line in Fig. 1b) arrives extremely late to the other five regions compared to the dates implied by the data (blue error bars for Cyprus, southern France, Catalonia, Navarre, and central Portugal). This is not surprising because, as mentioned above, it is known archeologically[5,6] that the Neolithic spread not only along an inland route (red color in Fig. 1b) but also following a sea route (blue color in Fig. 1b) which was faster and involved longer jumps (i.e., displacements per generation)[4,16,17]. We can take this into account by introducing a second model, which we call the sea model, so that farmers disperse with longer jumps along the coast (see Supplementary Information, Sec. S3 for details). We are not aware of any ethnographic data useful to estimate their length but, as seen in Fig. 1b, displacements of 70 km per generation lead to realistic arrival times (blue line) to the regional cultures reached by the sea route (blue error bars). It is true that in the *western* Mediterranean, the distance traveled along the coast by early farmers was much longer than 70 km[4,16,17], so we admit that more complicated models could be built by assuming a longer jump distance for the western[16] than for the eastern[3] Mediterranean and/or other local features. However, such complications are unlikely to change our conclusions below, because the latter follow from an observed genetic cline at the continental level (not in a specific region such as, e.g., the western Mediterranean), and moreover, the percentage of haplogroup K changes very slowly after the arrival of the first farmers (Supplementary Information, Sec. S3-C and Supplementary Fig. 6c). For the sea model (blue color in Fig. 1b), the distance is measured along the coast (see the white line in Fig. 1a and Supplementary Information, Sec. S3-B).

According to Fig. 1b, the average spread rate (at the continental scale) along the sea route was ~1.6 km/yr, almost twice that along the inland route (~0.9 km/yr). This difference and that seen above in the jump length (70 km along the sea route versus 50 km along the inland route) imply that geographical features had substantial effects on the dispersal behavior of early farmers. Moreover, the confidence intervals for the spread rates along both routes (estimated by linear regression of the archeological data shown as error bars in Fig. 1b) do not overlap with 95% confidence level (Supplementary Information, Sec. S7), and this confirms that there is a statistically significant difference between both dispersal rates.
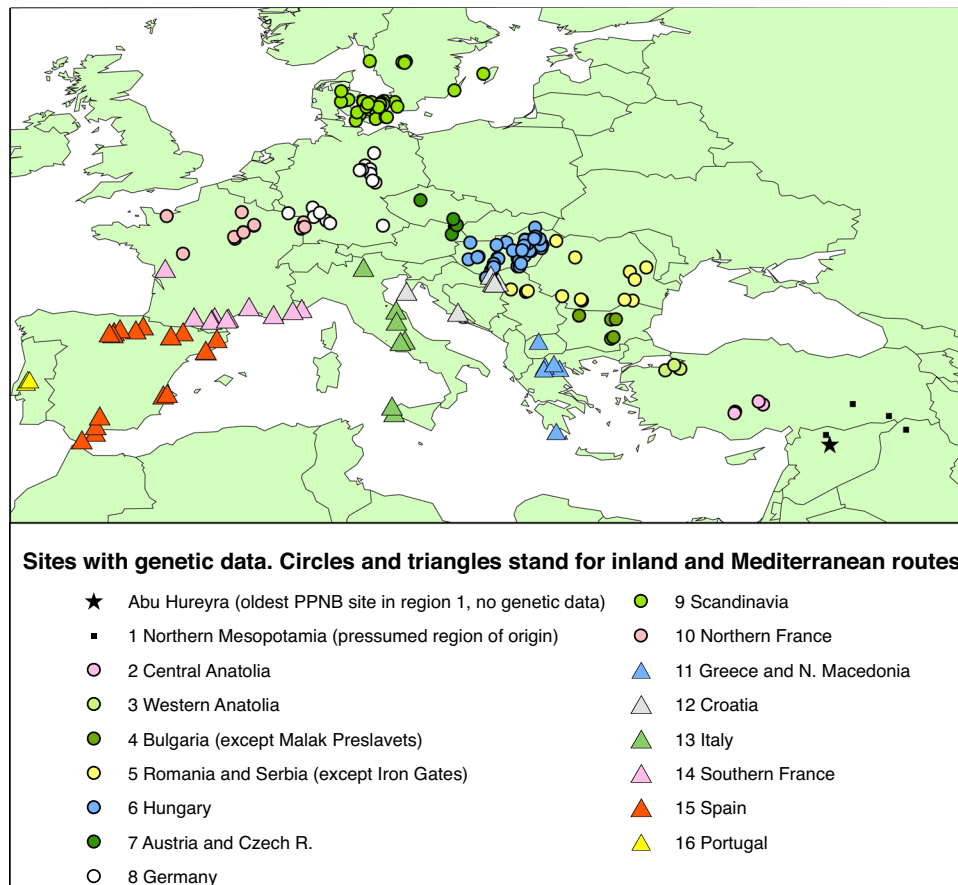
Fig. 2 | Sites with early Neolithic individuals whose mt haplogroup is known. Sites are grouped by the regions used for the inland route (circles) and the Mediterranean route (triangles) in Figs. 3 and 4. The squares (lower right) are sites in northern Mesopotamia (i.e., northern Syria, northwestern Iraq, and southeastern Anatolia). This is the presumed region from where the Neolithic spread across Anatolia (modern-day Turkey) and Europe.

A map of sites with early Neolithic genetic data is shown in Fig. 2. Due to the fact that the availability of genetic data depends on the region considered, in Figs. 2–4 regions are identified with numbers and some regions are different from those in Fig. 1, which are identified with letters. Since Fig. 1b yields reasonable agreement between our spatial simulations and the archeological data, it makes sense to compare our simulations also to genetic data. This is done in Fig. 3a for the inland route, and in Fig. 3b for the sea route. In Fig. 3 the results of our simulations are shown as lines, and the observed percentages of haplogroup K as error bars (from Supplementary Data 3). The error bars in Fig. 3 confirm conclusively the existence of a cline of mt haplogroup K for Neolithic populations, both along the inland route (Fig. 3a) and along the sea route (Fig. 3b), with decreasing values westwards and northwards. We have also confirmed formally the existence of both clines by means of spatial correlograms and the Bonferroni technique (Supplementary Information, Sec. S8).

We have seen (Fig. 1b and Supplementary Information, Sec. S7) that the average spread rate along the Mediterranean route (1.6 km/yr) was about twice faster than along the inland route (0.9 km/yr), and our simulations show that this difference can be explained if the dispersal distance per generation was substantially longer along the sea route (70 km) than inland (50 km). On the other hand, according to Fig. 3 the average slope of the genetic cline of haplogroup K along the Mediterranean route (-4×10^{-3} %/km) was half than that along the inland route (-8×10^{-3} %/km). Again, according to the simulations in Fig. 3 this difference can also be explained by the longer characteristic dispersal distance per generation along the Mediterranean route (70 km) than

along the inland route (50 km). The intuitive reason is that the former distance (70 km) is longer and, therefore, leads to fewer interbreeding events per unit distance (and thus to a lower decrease in the % of haplogroup K). Thus a single factor (namely, the difference in dispersal distance) explains both the differences in spread rate and in average slope of the genetic cline between both routes.

As expected from our Introduction, the steepness of the simulated cline (lines in Fig. 3a, b) increases with the intensity η of interbreeding (and/or acculturation). It is very interesting that, in spite of the difference in dispersal distance between the inland and Mediterranean routes, both genetic clines (Fig. 3a, b) imply a strikingly similar interaction behavior between farmers and HGs for both routes, because $\eta \approx 0.07$ in both cases. Indeed, the best overall agreement with the data is attained for the clines within the hatched areas (see the caption to Fig. 3), i.e., for $0.07 \leq \eta \leq 0.08$ in Fig. 3a (inland route) and $0.06 \leq \eta \leq 0.07$ in Fig. 3b (sea route). Both ranges are very similar and consistent with a single value ($\eta \approx 0.07$), in spite of the uncertainties in the data (error bars) and of the fact that the average slope of the inland cline is very different (twice larger) than that of the Mediterranean cline. We conclude that, whereas the dispersive behavior depended strongly on geography (i.e., the presence or not of the Mediterranean Sea), interbreeding and acculturation apparently did not depend on geographical constraints but only on the transition in the subsistence economy (from hunting and gathering to farming) and its associated way of life. Below we shall see that this conclusion is maintained when taking into account the uncertainties in the parameter values and the percentage of haplogroup K in the original population. The estimation
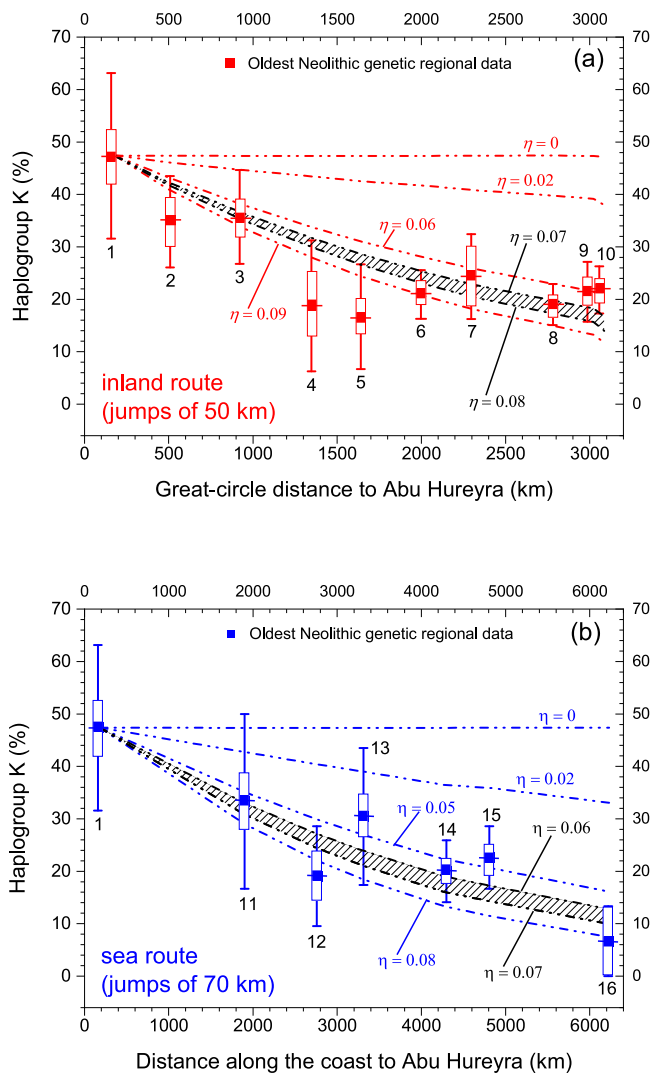
Fig. 3 | Observed (squares) and simulated (lines) percentages of mt haplogroup K among early farmers. a Inland route. **b** sea route. All error bars have been obtained from $n > 15$ individuals. Hatched areas indicate consistency between the observed data and the simulations, in the following sense. **a** The hatched area ($0.07 \leq \eta \leq 0.08$) is bounded by two black clines that cross all error bars except one. **b** The cline for $\eta = 0.06$ crosses all error bars and that for $\eta = 0.07$ crosses all error bars except one. **a** We have used, as in Fig. 1b, the inland model (jumps of 50 km) for regions located in the inland route, namely 1 northern Mesopotamia (northern Syria, northwestern Iraq and southeastern Anatolia, $n = 19$ individuals), 2 central Anatolia ($n = 46$), 3 Western Anatolia ($n = 56$), 4 Bulgaria (except Malak Preslavets, see Sec. S1-E, $n = 16$), 5 Romania and Serbia (except Iron Gates, see Sec. S1-E, $n = 30$), 6 Hungary ($n = 129$), 7 Austria and Czech Republic ($n = 37$), 8 Germany Linear-bandkeramik (LBK, $n = 179$), 9 Scandinavia ($n = 70$) and 10 northern France ($n = 133$). **b** We have used, as in Fig. 1b, the sea model (jumps of 70 km) for regions located along the Mediterranean route, namely 11 Greece and North Macedonia ($n = 18$), 12 Croatia ($n = 21$), 13 Italy ($n = 23$), 14 southern France ($n = 85$), 15 Spain ($n = 84$) and 16 Portugal ($n = 15$). The boxes (25% and 75% quartiles), error bars or whiskers (10% and 90% percentiles), and horizontal lines (medians) have been obtained by bootstrap resampling with replacement (10,000 replicates). Source data are provided in the excel Source Data file.

$\eta \approx 0.07$ corresponds to 3.6% of early farmers interbreeding with HGs and/or acculturating one of them (Methods, text below Eq. (14)).

The conclusion from Fig. 3 that $\eta \approx 0.07$ more than triplicates a previous result[20] which yielded $\eta \approx 0.02$. The main reason for this difference is the following methodological improvement introduced in this paper. The model in ref. 20 and some other previous approaches

(see Supplementary Information, Sec. S1-A) used real numbers for the population sizes, which obviously is not realistic because the number of individuals must necessarily be an integer (for details, see Supplementary Information, Sec. S2). Another difference with ref. 20 is that it did not analyze the inland and sea routes separately because the necessary genetic data were not available seven years ago (Methods).

In order to take the uncertainties in the parameter values into account, we computed envelopes on the simulation outputs. Figure 4 shows those for the inland route and leads to the range $0.06 \leq \eta \leq 0.12$, which refines that found in Fig. 3a without taking into account the uncertainties in the parameter values ($0.07 \leq \eta \leq 0.08$). We applied the same approach to the sea route and the result is $0.05 \leq \eta \leq 0.10$ (Supplementary Fig. 12b), which overlaps widely with that for the inland route, reinforcing our conclusion that the interaction between farmers and HGs along both routes was similar. The common range ($0.06 \leq \eta \leq 0.10$) implies that the percentage of farmers that interbred with HGs and/or acculturated them was 2.4%–5.9% (Methods, text below Eq. (14)).

Finally, we computed simulation envelopes similar to those in Fig. 4 but assuming that the percentage of haplogroup K in region 1 was not equal to the observed value (47.4% K, square 1 in Figs. 3 and 4) but to the lower (31.6% K) or upper (63.2% K) bound of the corresponding range (error bar 1 in Figs. 3 and 4). In all cases, there was wide overlap between both routes, and combining all possible cases leads to the overall range for the percentage of early farmers that interbred with HGs and/or acculturated them 1.2%–8.3% (Methods, text below Eq. (14)). We would like to stress that the inland and sea routes yield overlapping ranges of this percentage for any realistic set of parameter values and also for any realistic percentage of haplogroup K in the original population (Supplementary Information, Sec. S6).

Discussion

The first comparisons between human genetic clines along the inland (i.e., central European) and Mediterranean routes were performed more than 20 years ago using DNA from modern populations[11,44–46]. By then, it was not possible to compare observed clines to those obtained from simulations of the Neolithic wave of advance because ancient DNA was still unavailable, and the geographical patterns of modern DNA have been surely affected by population movements after the Neolithic. In contrast, here we have compared genetic clines along both routes using DNA obtained from ancient individuals. This has made it possible, for the first time to our knowledge, to estimate along each route (Fig. 3a, b) a key parameter describing the interaction behavior between early farmers and HGs, namely the intensity of interbreeding and/or acculturation η, and from this the percentage of farmers that interbred with HGs or acculturated one of them.

Some authors have previously estimated other interesting quantities that are related to the interaction between early farmers and HGs. One example is the HG ancestry proportion in early farmers[14,30,47], which is estimated using f_4 statistics (based on differences in allele frequencies)[48] and/or admixture analysis (which usually yields much the same results, see Fig. S6.2 in ref. 30). Other examples are the assimilation coefficient of HGs into farmer populations[49], the probability that an HG interbreeds with a farmer[50], etc. None of these results can be compared directly to our estimate η, because they measure different quantities. Two of these previous studies[49,50] modeled the Neolithic spread, but only along the inland route. Here we have compared the interaction behavior between farmers and HGs along the inland and Mediterranean routes. Another difference is that we have applied Eqs. (1)-(12), derived rigorously from cultural transmission theory[41,51].

It was previously known from archeological data[52–54] that the dispersive behavior of early European farmers depended on geography, because the Neolithic spread faster in coastal regions and this implies that early farmers moved longer distances per generation
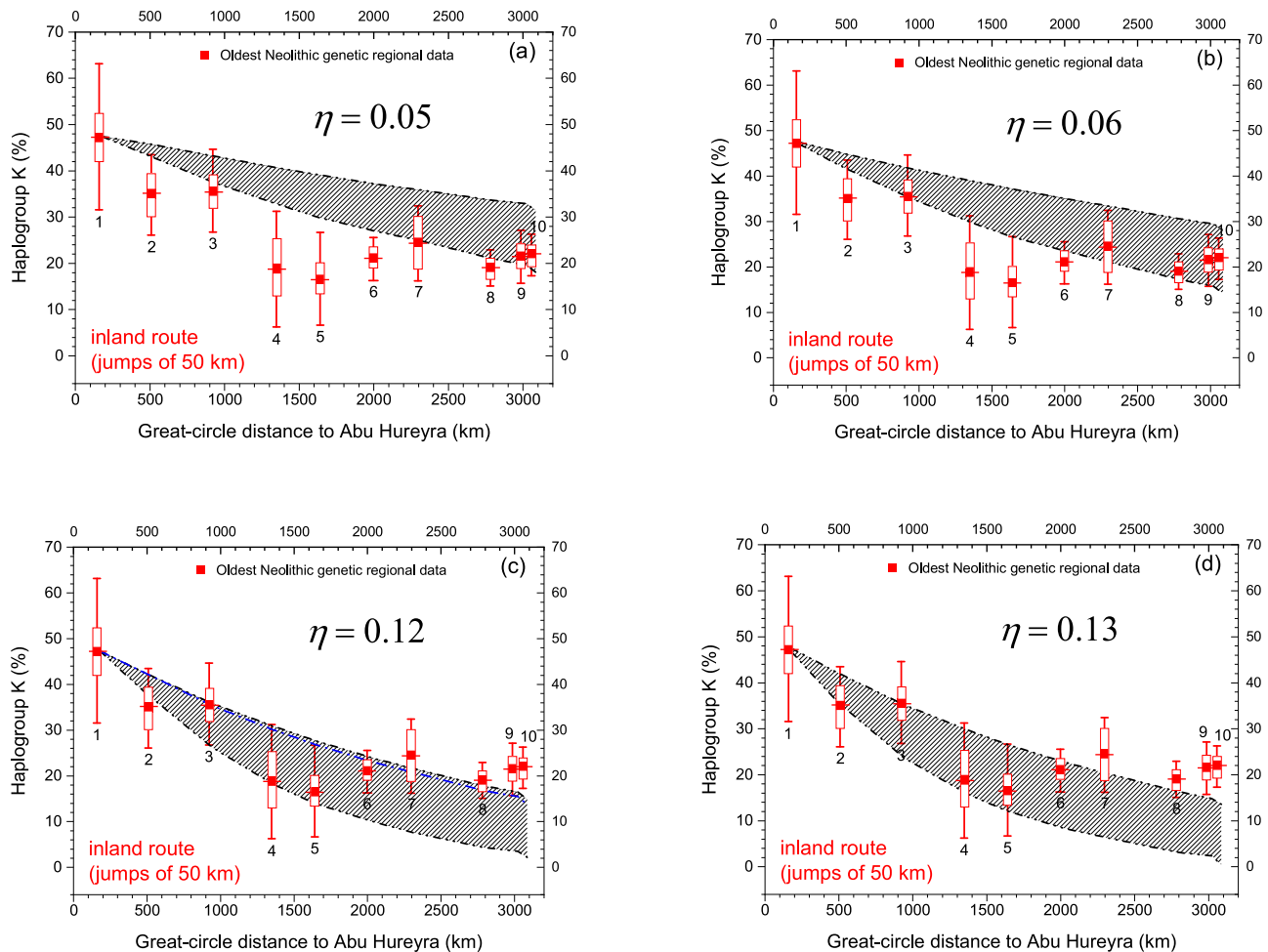
Fig. 4 | Observed and simulated percentages of mt haplogroup K (% K) among early farmers along the inland route. This figure takes into account the uncertainty in the parameter values, so each value of η has an envelope (rather than a line as in Fig. 3). For each distance, the hatched area gives the minimum and maximum values of the % K due to this uncertainty (see Methods). The data used are the same as in Fig. 3a, so the values of n, boxes, whiskers, and horizontal lines are also the same. This figure shows that no simulation outputs cross all error bars. Panel **a** ($\eta = 0.05$) shows that for $\eta \leq 0.05$ it is not possible to find any simulation output that crosses all error bars except one. In contrast, panel **b** shows that for $\eta = 0.06$ there is at least one simulation output (the lower black line) that crosses all error bars except one. Thus $\eta = 0.06$ is the minimum value of η consistent with the data. Panel **d** ($\eta = 0.13$) shows that for $\eta \geq 0.13$ it is not possible to find any simulation

output that crosses all error bars except one. In contrast, panel **c** ($\eta = 0.12$) shows that there is at least one simulation output that crosses all error bars except one (e.g., the blue line, see the caption to Supplementary Fig. 12a). Thus $\eta = 0.12$ is the maximum value of η consistent with the data. We conclude that assuming that the initial percentage of haplogroup K is equal to its observed value (47.4% K, square of error bar 1), consistency between the genetic data and the simulations in the inland route is possible only if $0.06 \leq \eta \leq 0.12$. This refines the range found in Fig. 3a without taking into account the uncertainties in the parameter values ($0.07 \leq \eta \leq 0.08$). Analogous figures for the sea route are included in Supplementary Fig. 12b and they lead to $0.05 \leq \eta \leq 0.10$, which refines the range from Fig. 3b ($0.06 \leq \eta \leq 0.07$). Source data are provided in the excel Source Data file.

along the Mediterranean coast than along the inland route[4,6,16,17]. Indeed, archeological data show that early migrating farmers moved on average ~50 km per generation along the inland route and ~70 km per generation along the Mediterranean route, with a spread rate (Fig. 1b and Sec. S7) in the latter case (~1.6 km/yr) almost twice than along the inland route (~0.9 km/yr). Thus geographical constraints had a strong effect on the dispersal behavior of early farmers in Europe. Here we have found that the slope of the genetic cline of haplogroup K along the Mediterranean route (~4 × 10^{-3} %/km) was half of that along the inland route (~8 × 10^{-3} %/km), again due to a strong geographical effect on human dispersal (Fig. 3). We have also shown the unexpected result that, in sharp contrast with these differences, the interactive behavior between farmers and HGs (interbreeding and/or acculturation) did not depend on geographical constraints (at the continental scale) because for both routes the percentage of farmers involved in this interaction was remarkably the same, namely ~3.6% or, more precisely, between 1.2% and 8.3%.

Our results provide a very simple explanation of the well-known higher HG ancestry along the Mediterranean route than along the inland one[31,32,40]. The sea route has ~6000 km along the coast (Fig. 3b), and along it early farmers moved ~70 km per generation on average (Fig. 1b), which implies ~6000/70 = 86 admixture events (one per generation). In contrast, along a straight line on the inland route, there were only ~3000/50 = 60 admixture events (Figs. 3a and 1b). So, given that the same percentage of farmers interbred with HGs in each admixture event along both routes (i.e., that the values of η in Fig. 3a, b are essentially the same), more HGs per farmer had been incorporated when the Neolithic wave of advance reached the end of the sea route that when it reached the end of the inland route. This explains that the proportion of HG ancestry was higher at the end of the sea route (Iberia) than at the end of the inland route (northern France), see e.g., Fig. 4a in ref. 31. This reasoning also explains three additional observations: (i) the HG ancestry proportion is generally higher in western European early farmers than in eastern ones[31,37]; (ii) at the end of the

sea route the percentage of haplogroup K is lower than at the end of the inland route (error bars in Fig. 3); and (iii) at the end of the sea route the percentage of HG haplogroups is higher than at the end of the inland route (Supplementary Data 7).

We caution that our results are valid in many regions (Fig. 2) but they do not exclude the possibility that the percentage of farmers that interbred with HGs and/or acculturated them deviated from our large-scale trend in specific locations or environments not included in our analysis. A case in point is that of some Danubian sites (such as Malak Preslavets in Bulgaria and Iron Gates sites in Serbia) with exceptional fishing resources and anomalously high HG famer interactions, as detected both archeologically and genetically[55]. Future work could address this topic explicitly, with more detailed simulations allowing for spatial variations in HG density, the proportion η of farmers that interbred with HGs or acculturated them, etc. Such non-homogeneous models would also be of interest to discuss scenarios in which admixture events do not take place in all locations but only in specific ones[11].

Our simple approach could be useful to simulate the *increase* in the percentages of HG haplogroups in the populations of early farmers along the inland and Mediterranean routes. We note that this is a substantially more complicated task, due to the need to know the precise initial genetic conditions (i.e., the detailed distribution of haplogroups in the populations of HGs in all regions analyzed, in addition to those of early farmers). Such an approach might help to clarify if some observed regional differences (e.g., the anomalously high HG haplogroup frequencies and HG ancestries observed in southern France[31,32,56]) are due to geographical differences in the initial genetic conditions of HGs, the initial densities of HGs, the population sizes of pioneering farmers[32] and/or key parameters of human behavior (such as the percentage of farmers that interbred with HGs and/or acculturated one of them).

It is worth stressing that the present study is based on a single marker (mt haplogroup K). We acknowledge that this makes our conclusions less certain than if we could check them using evidence from the complete genome. Admittedly, using data on a single marker implies limitations, in at least three different ways. Firstly, perhaps random effects were strong enough to have a major influence on the spatial distribution of haplogroup K. In this case, the clines that we have analyzed (Fig. 3) would not be due to interbreeding (as assumed by our model) but just to random effects. However, in Supplementary Information, Secs. S9, 10, we generate many clines at random and find that a spatial decrease similar to that of haplogroup K is not obtained (with 90% confidence level). This suggests that for haplogroup K the formation of the cline is driven by a non-random process that dominates over purely random effects. Our model includes such a process (interbreeding and/or acculturation), so it is reasonable to apply our model to analyze the clines of haplogroup K (other models are discussed and dismissed in Supplementary Information, Sec. S1-B).

A second possible limitation of our results is due to considering only one genetic marker, namely mt haplogroup K. There are two main motivations for us to use haplogroup K: (i) it is essentially absent in HGs, so we avoid the effect of the uncertainties in the regional frequencies of haplogroup K in HGs (Supplementary Information, Sec. S1-D). (ii) It is the mt haplogroup that reaches the highest regional frequency, ~50% (Supplementary Data 7), so effects due to the smallness of the regional samples available are less important that in lower-frequency haplogroups. Indeed, in Supplementary Information, Sec. S11, we show that, due to the smallness of the regional samples available at present, we can only detect clear clines for markers that display high frequencies. According to the data, there is not any haplogroup except K with a high frequency (~40% or more) in any region (Supplementary Data 7), so we have to use haplogroup K because all other haplogroups display too low frequencies (below 20%, see Supplementary Data 7b, c).
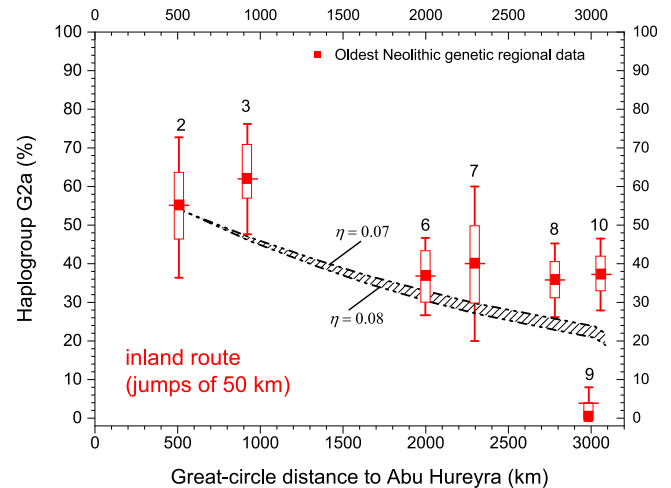

Fig. 5 | Observed (squares) and simulated (curves) percentages of Y-chromosome haplogroup G2a among early farmers along the inland route. The regions are the same as in Fig. 3. In order to avoid very large error bars, we have included only regions with $n>10$ individuals. The regions are 2 central Anatolia ($n = 11$), 3 Western Anatolia ($n = 21$), 6 Hungary ($n = 30$), 7 Austria and Czech Republic ($n = 10$), 8 Germany Linearbandkeramik (LBK, $n = 41$), 9 Scandinavia ($n = 20$) and 10 northern France ($n = 43$). (Supplementary Data 8). The boxes (25% and 75% quartiles), error bars or whiskers (10% and 90% percentiles) and horizontal lines (medians) have been obtained by bootstrap resampling with replacement (10,000 replicates). In region 9 there is not any individual with Y-chromosome haplogroup G2a, so we have applied the method in[20], Text S10. Source data are provided in the excel Source Data file.

A third possible limitation of our results is due to considering only mtDNA. It could be argued that analyzing other parts of the genome might, in principle, yield estimations for the intensity of interbreeding η inconsistent with the range obtained by us. Testing this possibility is difficult with the data available at present. Fortunately the Y-chromosome offers such a possibility, but there are much fewer data than for mtDNA (e.g., for region 1 we know the Y-chromosome haplogroup of only 2 individuals). In Fig. 5 we plot for the inland route the frequencies of haplogroup G2a, which was absent in HGs and is the Y-chromosome haplogroup that reaches the highest frequencies, above 50% (Supplementary Data 8). It exhibits a decreasing cline with increasing distance, as predicted by our model. However, it is true that the fact that the cline of G2a decreases implies only qualitative agreement with our model. Quantitative agreement requires that the observed cline is similar to those obtained from our simulations using the same parameter values as for the cline of mt haplogroup K (Fig. 3a). In order to test this, since our model can be directly applied to Y-chromosome DNA (as justified in Supplementary Information, Sec. S12), we use our model to obtain in Fig. 5 simulated clines using the values of η ($\eta = 0.07 - 0.08$) obtained in Fig. 3a from the cline of mt haplogroup K (all other parameter values are also the same as in Fig. 3a). In Fig. 5 there is reasonable agreement between the observed and simulated clines of haplogroup G2a. Thus Y-chromosome data, which are independent from mtDNA data, give support for our proposal. Unfortunately, we cannot perform an analogous analysis along the sea route, because at present only two regions have more than 10 individuals whose Y-chromosome haplogroup is known (Supplementary Data 8). In the future, the Y-haplogroup of more early farmers will be known and it will be possible to check: (i) if the percentage of G2a decreases also along the sea route, as predicted by our interbreeding model, and (ii) in case it does, if the corresponding interbreeding intensity η is consistent with the estimations in the present paper. We believe that our simple model cannot be applied directly to all parts of

the genome because, in general, we cannot define haplogroups in the sense used in the present paper due to recombination (exchange of genetic material between maternal and paternal chromosomes). However, when more data are available, additional tests of our proposal could be carried out for regions of autosomes that do not recombine or have essentially no recombination, see e.g.[57,58]. Moreover, as mentioned above, spatial clines are observed for some quantities that are calculated using the whole genome or a substantial part of it, e.g., the proportion of HG ancestry[31,32,37,59], diversity statistics[49], etc. Using sufficiently powerful computing equipment, our methods could be applied to simulate such clines. Comparing them to the observed ones could lead to estimations of the interbreeding/acculturation intensity η based on the whole genome, which could be compared to our estimations.

Methods
Observed data
We have gathered a database with the haplogroups of all early farmers from regions across which Archeology has shown that the Neolithic spread, from northern Mesopotamia to western and northern Europe. Recent ancient DNA data for northern Mesopotamia[60,61], central Anatolia[62–64], western Anatolia[60,65], Bulgaria[55,60], Scandinavia[66,67], northern France[31,68], Greece[49,55], Italy[37,69], southern France[31,68], Romania and Serbia[55,60] have made it possible to include also these regions in the present work (in contrast to ref. 20), so have now been able to analyze the genetic data of the inland and sea routes separately (Fig. 3a, b). Supplementary Data 1 is the Neolithic genetic database with all 961 early farmers along the inland and Mediterranean routes whose mt haplogroup is known. As already done in ref. 20, some regions have not been included in our analysis for reasons discussed in detail in Supplementary Information, Sec. S1-E (e.g., the southern Levant was excluded because its populations were not involved in the spread of the Neolithic across Anatolia and Europe[70]).

At a given location, the time of arrival of the first farmers does not necessarily coincide with the time estimated from the genetic material, because the individuals whose DNA has been recovered might have lived generations after the arrival of the first farmers. Thus, in order to develop a realistic spatial model, in addition to the dates of the individuals whose DNA has been retrieved, we also need the time of arrival of the Neolithic at each region, which is estimated from archeological (not genetic) data. All samples considered by us are either cereal seeds or bones from humans or domesticated animals, thus highly reliable. The date and location of the oldest archeological site for each of the regions listed in the caption to Fig. 1 is included as Supplementary Data 4.

Space-time genetic simulations
The spatial domain considered in our simulations encompasses Europe, the Near East and part of Asia and Africa. This area is represented by a grid of 300×300 squared cells of $50\,\text{km} \times 50\,\text{km}$ each in our simulations of the Neolithic spread to regions reached along the inland route (red color in Fig. 1b). The value 50 km has been estimated previously from ethnographic data of mobility per generation of pre-industrial farmers[42]. In contrast, for regions reached following the sea route we use a different simulation grid of cells with sides of 70 km, as required to reproduce the corresponding archeological arrival times (blue error bars in Fig. 1b). In a previous model[20] we used a non-homogeneous geography (with seas and mountains) whereas here we use an homogeneous one for 3 reasons: (i) simulations are faster and simpler; (ii) an homogeneous geography leads to very similar results than a non-homogeneous one (Supplementary Information, Sec. S2-A1); and (iii) an homogeneous geography makes it possible to test the models by comparing to analytical equations for the Neolithic spread rate (Supplementary Information, Sec. S2-A2). Another difference is that in ref. 20. the population density (and therefore the population

size) was represented using real numbers. This corresponds to the dispersal, reproduction and interaction of a non-integer number of individuals (e.g., less than one), which is obviously impossible. The present paper is based on a more realistic approach, in which population sizes have integer numbers (see below).

Each cell can host three populations: farmers with haplogroup K (with population size P_N), farmers without haplogroup K (with population size P_X) and HGs (with population size P_{HG}), where the population size is the number of individuals at a given spatial cell and time. For simplicity, HGs do not have haplogroup K, but modifying the simulations to include 2% or less of HGs with haplogroup K, as implied by the genetic evidence (Supplementary Data 5), leads to essentially the same results (Supplementary Fig. 1a).

Initial conditions
In order to compare the archeological data (error bars in Fig. 1b) to the simulations (full lines), it is necessary to assume an origin of distances (horizontal axis). In our simulations, the spatial origin corresponds to the place from where the Neolithic wave of advance spread into Anatolia and Europe. As a reasonable origin, we have used the site of Abu Hureyra (black square in Fig. 1a and star in Fig. 2) because of the following considerations. Neolithic traits (i.e., domesticated plants and animals) appeared at different places and times in the Near East, over several thousand years. Eventually, in the so-called pre-pottery Neolithic B and C (PPNB/C) cultures, a more homogeneous set of farming and stockbreeding practices (called the Neolithic package) was formed. It is from the PPNB/C cultures that the spread of the Neolithic into Europe proceeded[4]. According to the database in ref. 3, Abu Hureyra is the oldest PPNB site in Syria, Anatolia, and Iraq. Moreover, both spatial analysis of archeological dates (Fig. 3B in ref. 3) and recent genomic data[60,61] suggest that the region where Abu Hureyra is located (northern Mesopotamia, see Fig. 2 and Supplementary Fig. 1b) is likely the area from which the Neolithic spread into Anatolia and Europe. Furthermore, Abu Hureyra displays an exceptional, continuous record from foraging to farming, with sedentary HGs cultivating some cereals several thousand years before the Neolithic (i.e., before their economy became based on the cultivation of domesticated crops, rather than on hunting and gathering)[71]. It is also remarkable that the dietary shift in Abu Hureyra at the start of the Neolithic has been proven by dental evidence[72]. Hence not only radiocarbon dating, but also geostatistical, genomic, and archeological reasons support the use of Abu Hureyra as a reasonable origin of distances (as done in Figs. 1b and 3–5). According to Supplementary Data 4, the oldest reliable Neolithic date for Abu Hureyra is 9557 calibrated (cal.) years before the common era (yr BCE). This implies that the Neolithic front could have spread from there later but not sooner than at 9557 cal. yr BCE. In order to attain reasonable agreement between the simulated dates for the inland route (red line in Fig. 1b) and the archeological dates for the same route (red error bars in Fig. 1b), we have assumed that the Neolithic front spread from Abu Hureyra at 8718 cal. yr BCE. Such a date for the start of the Neolithic spread is within the PPNB period of Abu Hureyra[73]. In fact, the ultimate reason to use the initial time 8718 cal. yr BCE in our simulations is that we do not choose the slope of the red line in Fig. 1b because it is fixed by the characteristic dispersal distance (50 km), net fecundity, and generation time (given below), all of them obtained from ethnographic data.

Concerning the sea route, unfortunately we do not have an ethnographic value for a characteristic dispersal distance of pre-industrial farmer populations *living along a coast*, so the slope of the blue line in Fig. 1b (and thus the simulated spread rate along the *sea* route) is not fixed by ethnographic data but corresponds to choosing a dispersal distance of 70 km to obtain agreement with the blue error bars in Fig. 1b (see Results).

We assume that initially there are farmers only at the cell containing Abu Hureyra, in which we set the initial farmer population size

at its maximum possible value $P_{F\max} = l^2 p_{F\max}$, where l is the length of each side of the cell, i.e., the mean distance moved by generation (as used in Fig. 1b, l = 50 km for the inland route and l = 70 km for the sea route). The carrying capacity of farmers is $p_{F\max}$ = 1.28 individuals/km², a value from ethnographic data that has been used in previous genetic and archeological simulations[20,74]. For the entire grid (except the starting cell), we set the initial population of HGs equal to its saturation value, $P_{HG\max} = l^2 p_{HG\max}$ with $p_{HG\max}$ = 0.064 individuals/km², again from ethnographic data and as used previously[12,20,75]. Other realistic values of $P_{F\max}$ and $P_{HG\max}$ do not change the conclusions (see the sensitivity analyses in Supplementary Information, Secs. S5 and S6).

We use the simulations to predict the percentage of haplogroup K at several locations (their latitudes and longitudes are transformed into X and Y coordinates on the simulation grid by following the approach explained in Supplementary Information, Sec. S3). As explained above, for the oldest PPNB *archeological* site in northern Mesopotamia (Abu Hureyra) we use the date 8718 yr BCE. However, the average date of the early farmers from northern Mesopotamia for which we have *genetic* data is more recent, namely 7752 yr BCE, and their average location is also different from that of Abu Hureyra (Supplementary Data 1). In order to take into account this difference in dates and locations, we first found (by trial and error) the percentage of haplogroup K (% K) that needs to be assumed at the initial time (8718 yr BCE) and origin (Abu Hureyra) of the dispersal in order to obtain the observed % K (47.4%, from Data S3) at the average time (7752 yr BCE) and location of the early farmers from northern Mesopotamia for which we have *genetic* data.

Population dynamics

Front propagation models have been implemented to analyze human dispersals[41]. In such phenomena, changes in population numbers result from two processes, namely net population growth (reproduction minus deaths) and migration (dispersal). Our model also includes cultural transmission (i.e., interbreeding and/or acculturation). In the simulations each time step is one generation. The generation time, defined as the average age of a parent at the time when one of his/her children is born, is T = 32 yr and has been estimated using observed data for pre-industrial farming populations[76]. For each generation, the model proceeds sequentially through the following three steps.

Dispersal. At each inland cell, the model computes the number of farmers P_N and P_X (i.e., with and without haplogroup K, respectively) arriving from each of the four adjacent cells. Although several dispersal distances could be included, we expect that the results would be similar to those from our simple isotropic single-distance dispersal model, already applied in previous work[20,42]. Similarly, at each coastal cell (grid edge), individuals can arrive from the adjacent inland cell (in the direction orthogonal to the coast) and from the two adjacent coast cells. Ethnographic data indicate that a fraction of about p_e = 0.38 of farmers (called the persistence) stays at each cell[42].

In principle, the easiest approach to implement integer numbers could seem to assume that the number of farmers (of each group N and X) that stays at each cell is equal to the nearest integer to the corresponding initial number multiplied by the persistence p_e. Similarly, we could be tempted to assume that the number of individuals who jump to each of the 4 first neighboring cells is the nearest integer to the initial number multiplied by $(1 - p_e)/4$. However, this simple approach has a serious problem. Assume, for example, that a node has P_N = 3 and P_X = 5 farmers. Then the number of N-individuals that jump in each of the 4 directions would be $\text{NINT}((1 - 0.38)3/4) = 0$, and the corresponding number of X-individuals would be $\text{NINT}((1 - 0.38)5/4) = 1$. Thus, in the nodes where farmers arrive, their percentage with haplogroup K would be 0%. This effect is due to the smallness of the population size (obviously, the result 0% would not be obtained if, instead of P_N = 3 and P_X = 5, we considered e.g. P_N = 30 and P_X = 50). In

fact, this effect disappears if we realize that such a simple approach is unrealistic because both ethnographic[77] and archeological[5,78] data imply that human populations do not live in groups of arbitrarily low size. As explained in detail in Supplementary Information, Secs. S2-B–D, we take this into account by imposing the condition that dispersal takes place only after the number of farmers in a given node reaches a minimal number (dispersal threshold). Previous simulations of Neolithic spread have already applied a dispersal threshold[9], which is very reasonable because there is a well-known minimum population size for human groups, which is thought to be related to the benefits of food-sharing, division of labor, and other forms of cooperation[77]. Indeed, archeologists have remarked that the first Neolithic settlements in local regions generally consisted of several houses, not a single one (see, e.g., ref. 5, pp. 97). We use a minimum threshold for the population size corresponding to a density of $p_{F\min}$ = 0.06 individuals/km² according to Early Neolithic archeological data[78] (see Supplementary Information, Sec. S2C). In our model the number of individuals is conserved by using a random dispersal algorithm and requiring that the number of individuals that stay in each cell is equal to the initial number (before dispersal) minus the sum of the numbers that jump in the four directions (for details see Supplementary Information, Sec. S2-D). Thus, the model is stochastic, and the number of individuals of any type that jump into the four cells is not necessarily the same. When the first farmers arrive at a cell, the number of HGs begins to decrease due to cultural transmission (see the next paragraph), and we can neglect the dispersal of HGs because their decrease is compensated by some farmers and their offspring who occupy the space formerly used only by HGs.

Cultural transmission. Culture can be transmitted between generations by three different mechanisms[51]. If farmers and HGs interbreed, this transmission of culture is called vertical and it is well-known from ethnographic fieldwork that the children of these mixed couples are farmers[29,79]. In contrast, acculturation occurs when HGs learn agriculture from farmers of the same generation (horizontal transmission) or the previous one (oblique transmission). Out of the three mechanisms mentioned, we only apply vertical transmission (i.e., interbreeding), but including horizontal/oblique transmission (i.e., acculturation) in addition to (or in place of) vertical transmission would lead to the same conclusions, with the only difference that η would not be the intensity of interbreeding but of interbreeding and acculturation (see Text S9 in ref. 20).

The number of cross-matings (i.e., mixed couples) between HGs (H) and each of the two populations of farmers (N and X) in a cell of the simulation grid is given by equations derived using cultural transmission theory[51] in ref. 80, namely (see also Text S5 in ref. 20)

$$couples\ HN = \eta \frac{P_{HG} P_N}{P_{HG} + P_N + P_X}, \qquad (1)$$

$$couples\ HX = \eta \frac{P_{HG} P_X}{P_{HG} + P_N + P_X}, \qquad (2)$$

where the denominators (i.e., $P_{HG} + P_N + P_X$) give the total population at the cell considered, and parameter η is called the intensity of interbreeding ($0 \le \eta \le 1$)[80]. The number of individuals who do not take part in HN or HX matings is obviously given by

$$P'_{HG}(x,y,t) = P_{HG}(x,y,t) - couples\ HN - couples\ HX, \qquad (3)$$

$$P'_N(x,y,t) = P_N(x,y,t) - couples\ HN, \qquad (4)$$

$$P'_X(x,y,t) = P_X(x,y,t) - couples\ HX, \qquad (5)$$

where the number of couples is given by Eqs. (1)-(2). Since there is no reason to assume that farmers of a genetic group (K or non-K) have a preference for (neither against) mating with farmers of the same genetic group, we assume random mating among farmers (η = 1)[51,80], i.e.

$$\text{couples } NX = \frac{P'_N P'_X}{P'_N + P'_X}. \tag{6}$$

Reproduction. The net fecundity is defined as the number of children per parent that survive and reproduce (because obviously, only these affect the propagation of the Neolithic front). Let $R_{0,HG}$ stand for the net fecundity of HGs (which applies to matings in which both parents are HGs) and $R_{0,F}$ for the net fecundity of farmers (which applies to all other matings because, as mentioned above, the children of cross-matings between farmers and HGs are farmers). Next, we apply Eqs. (1)-(6) to find out the number of individuals for each population (P_{HG}, P_N and P_X) at generation $t+1$ as a function of those at generation t. The simplest case is that of HGs, because both parents of a HG are necessarily HGs. Thus

$$P_{HG}(x,y,t+1) = R_{0,HG}\left[2\,couples\,HH\right], \tag{7}$$

where we have taken into account that $R_{0,HG}$ is the net fecundity (number of surviving children) per person, so the net fecundity per couple is $2R_{0,HG}$. In Eq. (7), $couples\,HH = P'_{HG}(x,y,t)/2$ is the number of couples in which both mates are HGs. Using Eq. (3), we obtain

$$P_{HG}(x,y,t+1) = R_{0,HG}\left[P_{HG}(x,y,t) - couples\,HN - couples\,HX\right], \tag{8}$$

where the last two terms are given by Eqs. (1)-(2). We next consider the case of farmers, which is less straightforward[20] because one of the parents of a farmer can be an HG (see point (ii) above) or a farmer of another genetic type. Assuming for simplicity that in half of the mixed genetic matings (HN and NX), the mother belongs to group N (farmers with haplogroup K) and taking into account that, since mtDNA is passed from mother to children, her children will also have haplogroup K, we can apply that 50% of the offspring of such matings will be of type N. Then it is obvious that

$$P_N(x,y,t+1) = R_{0,F}\left[2\,couples\,NN + couples\,HN + couples\,NX\right], \tag{9}$$

$$P_X(x,y,t+1) = R_{0,F}\left[2\,couples\,XX + couples\,HN + couples\,NX + 2\,couples\,HX\right]. \tag{10}$$

In the last term of Eq. (10) we have applied that HGs and farmers of group K lack haplogroup K, so all offspring of *couples HX* will also lack it, i.e., they will be all farmers of type X.

Substantially more complicated equations with six populations (to distinguish females from males) and taking into account that all HGs are women in cross-matings between HGs and farmers (as suggested by ethnographic data) lead to very similar results. This was explicitly shown in a model using real numbers for the population sizes (Text S11 in ref. 20) and we expect that the same will happen using integer numbers instead.

The number of couples of farmers that mate with farmers of the same genetic type are those that do not mate with HGs or farmers of the other genetic type, i.e. *couples NN* = $\left[P'_N(x,y,t) - couples\,NX\right]/2$ and *couples XX* = $\left[P'_X(x,y,t) - couples\,NX\right]/2$. Using these numbers of couples into Eqs. (9), (10), and then Eqs. (4), (5) we finally obtain

$$P_N(x,y,t+1) = R_{0,F}P_N(x,y,t), \tag{11}$$

$$P_X(x,y,t+1) = R_{0,F}\left[P_X(x,y,t) + couples\,HN + couples\,HX\right]. \tag{12}$$

At each cell and generation, the effect of cultural transmission is computed using Eqs. (8), (11) and (12). For clarity, we have presented our derivation in the context of the Neolithic transition, but these equations are valid in general (i.e., for any population, epoch, and cultural trait), with suitable modifications if necessary. In accordance with ethnographic data, we set the net fecundity of farmers equal to $R_{0,F} = 2.45$ as in previous work[20,42,81] which means that at each node, the new population is generated by multiplying 2.45 times the size of the parent population. The resulting number is adjusted to the nearest integer (see Supplementary Information, Sec. S2-B). A setback of this simple implementation is that it yields an infinite growth that should be limited by the carrying capacity of farmers, $P_{F\,max}$ (where $P_F = P_N + P_X$). Consequently, at a given cell, if the number of farmers is above its maximum, it is set to its maximum possible value ($P_{F\,max}$) without changing the relative fractions of farmers with and without haplogroup K. We expect that using a logistic model would not change the conclusions because both models are very similar except when P_F approaches $P_{F\,max}$ and it is well-known that this does not effect the front dynamics[42].

We assume that the HG populations at each cell are at a steady state ($R_{0,HG} = 1$) and at saturation ($P_{HG} = P_{HG\,max}$) before the arrival of farmers. After the latter arrive to a cell, its population size of HGs will decrease due to cultural transmission. In turn, this implies that some farmers and their offspring will be in the space formerly occupied by HGs, so we assume that the remaining HGs maintain their steady state, $R_{0,HG} = 1$ (as in previous work, which used real numbers[20], we expect that other reasonable values of $R_{0,HG}$ will lead to very similar results using integer numbers).

Our Fortran programs (publicly available at https://zenodo.org/records/11099220) apply, at each cell and time increase (corresponding to one generation), the three steps of the cycle explained above, i.e., (i) dispersal, (ii) cultural transmission, and (iii) reproduction. In this way, the program finds out the number of individuals of the three populations in each region at the cell corresponding to the average latitude and longitude over all its early farmers whose mitochondrial haplogroup is known, at the average date of the same individuals (Supplementary Data 1). From this, we know the regional fractions of farmers with haplogroup K, namely $\frac{P_N}{(P_N + P_X)}100$. The time span from the beginning of the Neolithic spread (Syria, 8718 cal. yr BCE) up to the average *genetic* date of the most recent region in the genetic database (Scandinavia, 3286 cal. yr BCE, see Supplementary Data 1) is 5432 yr. As mentioned above we set $T = 32$ yr for the time elapsed in one generation, so we run our model for $t = 200$ iterations to cover the necessary time span. As explained above, the dispersal step is stochastic, and the model, therefore, yields slightly different results in different simulation runs, even for the same set of parameter values. However, for any given values of the position and time, the differences in the percentage of haplogroup K between different simulation runs are -0.1% or less, so the clines (Fig. 3) do not change (i.e., it is not necessary to average over different runs). For detailed discussions on the model, see Supplementary Information, Secs. S2–S4.

The uncertainties in the parameter values are taken into account by considering realistic ranges from ethnographic data and analyzing their effect on the simulated cline (Supplementary Information, Secs. S5, S6). In this way it is found that, for a given location and value of η, the maximum percentage of haplogroup K (upper black lines in Fig. 4) is obtained for $p_{F\,min} = 0.07$ farmers/km², $R_{0,F} = 2.87$, $p_{F\,max} = 1.86$ farmers/km² and $p_{HG\,max} = 0.047$ HGs/km², and the minimum one (lower black lines in Fig. 4) for $p_{F\,min} = 0.05$ farmers/km², $R_{0,F} = 2.09$, $p_{F\,max} = 0.96$ farmers/km² and $p_{HG\,max} = 0.072$ HGs/km². The generation time T and persistence p_e have a negligible effect as long as ethnographically realistic values for them are used[42,76], and similarly the dispersal distance l is tightly constrained by the archeological data (Fig. 1b). Envelopes similar to Fig. 4 for the sea route and/or other values of the percentage of haplogroup K in the original population (error bar 1 in Figs. 3 and 4) are included in Sec. S6.

Finally we focus on cultural transmission by computing the relative increase of the total population of farmers $P_F = P_N + P_X$ during one generation leaving aside reproduction ($R_{0,F} = 1$) and dispersal from Eqs. (11), (12), (1) and (2),

$$\frac{P_F(x,y,t+1) - P_F(x,y,t)}{P_F(x,y,t)} = \eta \frac{P_{HG}(x,y,t)}{P_{HG}(x,y,t) + P_F(x,y,t)}. \tag{13}$$

The numbers of individuals per square cell with side l are related to the respective population densities as $P_{HG} = p_{HG} l^2$ and $P_F = p_F l^2$. As explained above, when the first farmers arrive to a cell $p_F = p_{F\min}$ and $p_{HG} = p_{HG\max}$. Thus the percentage of early farmers that interbreed with HGs (or acculturate them) is simply

$$\frac{100\eta}{1 + \frac{p_{F\min}}{p_{HG\max}}} \tag{14}$$

In the first part of these Methods we have seen that Fig. 3 has been obtained for $p_{F\min} = 0.06$ individuals/km^2 and $p_{HG\max} = 0.064$ individuals/km^2. Also in Fig. 3, the best agreement between data and simulations is attained for $\eta = 0.07$. Using these three values in Eq. (14) we estimate that the percentage of farmers that interbred with HGs or acculturated them is 3.6%.

The envelopes for the inland route (Fig. 4) and the sea route (Supplementary Fig. 12b) have been obtained for $0.05 \leq p_{F\min} \leq 0.07$ individuals/km^2 and $0.96 \leq p_{F\max} \leq 1.86$ individuals/km^2 (for the rest of parameter values, see Supplementary Information, Sec. S6-A). These envelopes agree with the data if $0.06 \leq \eta \leq 0.10$ (Supplementary Fig. 12a, b). Using these three ranges Eq. (14) leads to the estimation 2.4%–5.9% for the percentage of farmers that interbred with HGs or acculturated them. This refines the estimation of 3.6% obtained in the previous paragraph without taking into account the uncertainty in the parameter values.

The previous two paragraphs assume a percentage of haplogroup K in the original population equal to its observed value (47.4% K, square 1 in Figs. 3 and 4). If we assume, instead, that this percentage is equal to its lower bound (31.6% K, error bar 1 in Figs. 3, 4), the data and the simulations agree for both the inland and sea routes if $0.03 \leq \eta \leq 0.06$ (Supplementary Fig. 13a, b) and Eq. (14) leads to the range 1.2%–3.5% for the percentage of farmers that interbred with HGs or acculturated one of them. Finally, if the percentage of haplogroup K in the original population is equal to its upper bound (63.2% K, error bar 1 in Figs. 3, 4), agreement between data and simulations along the inland and sea routes implies that $0.10 \leq \eta \leq 0.14$ (Supplementary Fig. 14a, c) and Eq. (14) yields 4.0%–8.3% for the percentage of farmers that interbred with HGs or acculturated one of them. Therefore, the overall range is 1.2%–8.3% (as reported in the Results section).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used are included as a single excel file (Supplementary Data 1–8). Source data for figures are provided as a Source Data file. Source data are provided with this paper.

Code availability

All computer programs used and a README file are publicly available at https://zenodo.org/records/11099220.

References

1. Bocquet-Appel, J. P. & Bar-Yosef, O. The Neolithic demographic transition and its consequences. (Berlin: Springer, 2008).
2. Premack, D. & Premack, A. Evolution versus invention. *Science* **307**, 673 (2005).
3. Pinhasi, R., Fort, J. & Ammerman, A. J. Tracing the origin and spread of agriculture in Europe. *PLoS Biol.* **3**, e410 (2005).
4. Fort, J., Pujol, T. & vander Linden, M. Modelling the Neolithic transition in the Near East and Europe. *Am. Antiq.* **77**, 203–220 (2012).
5. Shennan, S. The first farmers of Europe. An evolutionary perspective, Cambridge: Cambridge University Press, 2018.
6. Zilhao, J. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *PNAS* **98**, 14180–14185 (2001).
7. Childe, V.G. The dawn of European civilization, London: Kegan Paul, 1925.
8. Edmonson, M. S. Neolithic diffusion rates. *Curr. Anthropol.* **2**, 74–102 (1961).
9. Barbujani, G., Sokal, R. R. & Oden, N. L. Indo-European origins: a computer-simulation test of five hypotheses. *Am. J. Phys. Anthropol.* **96**, 109–132 (1995).
10. Chikhi, L., Destro-Bisol, G., Bertorelle, G., Pascali, V. & Barbujani, G. Clines of nuclear DNA markers sugest a largely Neolithic ancestry of the European gene pool. *PNAS* **95**, 9053–9058 (1998).
11. Chikhi, L., Nichols, R. A., Barbujani, G. & Beaumont, M. A. Y genetic data support the Neolithic demic diffusion model. *PNAS* **99**, 11008–11103 (2002).
12. Currat, M. & Excoffier, L. The effect of the Neolithic expansion on European molecular diversity. *Proc. R. Soc. B* **272**, 679–688 (2005).
13. Rasteiro, R. & Chikhi, L. Female and male perspectives on the Neolithic transition in Europe: clues from ancient and modern genetic data. *PLoS One* **8**, e60944 (2013).
14. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
15. Nikitin, A. G. et al. Interactions between earliest Linearbandkeramik farmers and central European hunter-gatherers at the dawn of European Neolithization. *Sci. Rep.* **9**, 19544 (2019).
16. Isern, N., Zilhao, J., Fort, J. & Ammerman, A. J. Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. *PNAS* **114**, 897–902 (2017).
17. Fort, J. Dispersal distances and cultural effects in the spread of the Neolithic along the northern Mediterranean coast. *Archaeol. Anthropol. Sci.* **14**, 153 (2022).
18. Fort, J. Prehistoric spread rates and genetic clines. *Hum. Popul. Genet. Genom.* **2**, 0003 (2022).
19. Sgaramella-Zonta, L. & Cavalli-Sforza, L. L. Method for the detection of a genetic cline, in *Genetic structure of populations*, Honolulu, University of Hawaii Press, 128–135 (1973).
20. Isern, N., Fort, J. & Rioja, V. L. The ancient cline of haplogroup K implies that the Neolithic transition in Europe was mainly demic. *Sci. Rep.* **7**, 11229 (2017).
21. Bertranpetit, J. Genetics of population history. The case of the Iberian peninsula and the origin of Basques. *Hum. Popul. Genet. Genom.* **2**, 0002 (2021).
22. Cardoso, S. et al. The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-Neolithic genetic substrate of Basques. *PLoS One* **8**, e67835 (2013).
23. Olalde, I. et al. The genomic history of the Iberian peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
24. Bramanti, B. et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137–140 (2009).
25. Brandt, G. et al. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257–261 (2013).
26. Rohrlach, A. B. et al. Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe. *Sci. Rep.* **11**, 15005 (2021).

27. Brotherton, P. et al. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature Communications* **4**, 1764 (2013).

28. Nieves-Colón, M. A. et al. Ancient DNA reconstructs the genetic legacies of precontact Puerto Rico communities. *Mol. Biol. Evol.* **37**, 611–626 (2020).

29. Ammerman, A. J. & Cavalli-Sforza, L. L. The neolithic transition and the genetics of populations in Europe, New Jersey: Princeton University Press, 1984.

30. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).

31. Rivollat, M. et al. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci. Adv.* **6**, eaaz5344 (2020).

32. Arzelier, A. et al. Neolithic genomic data from southern France showcase intensified interactions with hunter-gatherer communities. *iScience* **25**, 105387 (2022).

33. Villalba-Mouco, V. et al. Survival of late pleistocene hunter-gatherer ancestry in the Iberian peninsula. *Curr. Biol.* **29**, 1–9 (2019).

34. Allentoft, M. E. et al. 100 ancient genomes show repeated population turnovers in Neolithic Denmark. *Nature* **625**, 329–337 (2024).

35. Olalde, I. et al. A common genetic origin for early farmers from Mediterranean Cardial and central European LBK cultures. *Mol. Biol. Evol.* **32**, 3132–3142 (2015).

36. Betti, L. et al. Climate shaped how Neolithic farmers and European hunter-gatherers interacted after a major slowdown from 6,100 BCE to 4,500 BCE. *Nat. Hum. Behav.* **4**, 1004–1010 (2020).

37. Yu, H. et al. Genomic and dietary discontinuities during the Mesolithic and Early Neolithic in Sicily. *iScience* **25**, 104244 (2022).

38. Valdiosera, C. et al. Four millenia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *PNAS* **115**, 3428–3433 (2018).

39. Ariano, B. et al. Ancient Maltese genomes and the genetic geography of Neolithic Europe. *Curr. Biol.* **32**, 1–13 (2022).

40. Cox, S. L. et al. Socio-cultural practices may have affected sex differences in stature in early Neolithic Europe. *Nat. Hum. Behav.* **8**, 243–255 (2024).

41. Fort, J. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. *Proc. Natl. Acad. Sci.* **109**, 18669–18673 (2012).

42. Fort, J., Pérez-Losada, J. & Isern, N. Fronts from integrodifference equations and persistence effects on the Neolithic transition. *Phys. Rev. E* **76**, 031913 (2007).

43. Pérez-Losada, J. & Fort, J. Age-dependent mortality, fecundity and mobility effects on front speeds: theory and application to the Neolithic transition. *J. Stat. Mechs. Theor. Exp.* **2010**, P11006 (2010).

44. Semino, O. et al. The genetic legacy of Paleolithic Homo Sapiens: a Y chromosome perspective. *Science* **290**, 1155–1159 (2000).

45. Simoni, L. et al. Geographic patterns of mtDNA diversity in Europe. *Am. J. Hum. Genet.* **66**, 262–278 (2000).

46. Simoni, L. et al. Spatial variation of mtDNA hypervariable region I among European populations,. In: *Archaeogenetics: DNA and the population prehistory of Europe,* C. Renfrew and K. Boyle, Eds., Cambridge, McDonald Institute for Archaeological Research, 2000, 131-138.

47. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).

48. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

49. Silva, N. K. et al. Ancient mitochondrial diversity reveals population homogeneity in Neolithic Greece and identifies population dynamics along the Danubian expansion axis. *Sci. Rep.* **12**, 13474 (2022).

50. Marchi, N. et al. The mixed genetic origin of the first farmers of Europe. *Cell* **184**, 1842–1859 (2022).

51. Cavalli-Sforza, L. L. & Feldman, M. W. Cultural transmission and evolution: a quantitative approach, New Jersey: Princeton University Press, 1981.

52. Ammerman, A. J. & Cavalli-Sforza, L. L. Measuring the rate of spread of early farming in Europe. *Royal Anthropological Institute of Great Britain and Ireland* **6**, 674–688 (1971).

53. Bocquet-Appel, J.-P., Naji, S., vander Linden, M. & Kozlowski, J. Understanding the rates of expansion of the farming system in Europe. *J. Arch. Sci.* **39**, 531–546 (2012).

54. Henderson, D. A. et al. Regional variations in the European Neolithic dispersal: the role of the coastlines. *Antiquity* **88**, 1291–1302 (2014).

55. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).

56. Rivollat, M. et al. When the waves of European neolithization met: first paleogenetic evidence from early farmers in the Southern paris basin. *PLoS One* **10**, e0125521 (2015).

57. Jin, L. et al. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *PNAS* **96**, 3796–3800 (1999).

58. Tishkoff, S. A. et al. Global patterns of linkage disequilibrium at the CD4 locus and modern human origin. *Science* **271**, 1380–1387 (1996).

59. Racimo, F. et al. «The spatiotemporal spread of human migrations during the European Holocene,». *PNAS* **117**, 8989–9000 (2020).

60. Lazaridis, I. et al. Ancient DNA from mesopotamia suggests distinct pre-pottery and pottery neolithic migrations into Anatolia. *Science* **377**, 982–987 (2022).

61. Altınışık, N. E. et al. A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic Transition. *Sci. Adv.* **8**, eabo3609 (2022).

62. Feldman, M. et al. Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat. Commun.* **10**, 1218 (2019).

63. Chyleński, M. et al. Ancient mitochondrial genomes reveal the absence of maternal kinship in the burials of Çatalhöyük people and their genetic affinities. *Genes* **10**, 207 (2019).

64. Yaka, R. et al. Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. *Curr. Biol.* **31**, 1–14 (2021).

65. Blöcher, J. Genetic variation related to the adaptation of humans to an agriculturalist lifestyle. *Dissertation.* Johannes Gutenberg-Universität, Mainz (2019).

66. Blank, M. et al. Mobility patterns in inland southwestern Sweden during the Neolithic and Early Bronze Age. *Archaeol. Anthropol. Sci.* **13**, 64 (2021).

67. Allentoft, M. E. et al. Population genomics of post-glacial western Eurasia. *Nature* **625**, 301–311 (2024).

68. Brunel, S. et al. Ancient genomes from present-day France unveil 7,000 years of its demographic history. *Proc. Natl. Acad. Sci.* **117**, 12791–12798 (2020).

69. Antonio, M. L. et al. Ancient Rome: a genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714 (2019).

70. Lazardis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).

71. Hillman, G., Hedges, R., Moore, A., Colledge, S. & Pettitt, P. New evidence of Lateglacial cereal cultivation at Abu Hureyra on the Euphrates. *Holocene* **11**, 383–393 (2001).

72. Molleson, T. & Jones, K. Dental evidence for dietary change at Abu Hureyra. *J. Arch. Sci.* **18**, 525–539 (1991).

73. Smith, A., Oechsner, A., Rowley-Cowny, P. & Moore, A. M. T. Epipalaeolithic animal tending to Neolithic herding at Abu Hureyra, Syria (12,800-7,800 cal BP): deciphering dung spherulites. *PLoS One* **17**, e0272947 (2022).

74. Currat, M. & Excoffier, L. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biology* **2**, 2264–2274 (2002).

75. Fort, J., Pérez-Losada, J., Suñol, J. J., Escoda, L. & Massaneda, J. M. Integro-difference equations for interacting species and the Neolithic transition. *New J. Phys* **10**, 043045 (2008).

76. Fort, J., Jana, D. & Humet, J. M. Multidelayed random walks: theory and application to the Neolithic transition in Europe. *Phys. Rev. E* **70**, 031913 (2004).

77. Wobst, M. Boundary contitions for Paleolithic social systems: a simulation approach. *Am. Antiq.* **39**, 147–178 (1974).

78. Zimmerman, A., Hilpert, J. & Wendt, K. P. Estimations of population density for selected periods between the Neolithic and AD 1800. *Hum. Biol.* **81**, 357–380 (2009).

79. Cronk, L. From hunters to herders: subsistence change as a reproductive strategy among the Mukogodo. *Curr. Anthropol.* **30**, 224–234 (1989).

80. Fort, J. Vertical cultural transmission effects on demic front propagation: theory and application to the Neolithic transition in Europe. *Phys. Rev. E* **056124**, 1–10 (2011).

81. Isern, N., Fort, J. & Pérez-Losada, J. Realistic dispersion kernels applied to cohabitation reaction-dispersion equations. *J. Stat. Mech. Theor. Exp.* **10**, P10012 (2008).

Author contributions

J.F devised the project, prepared the final database and figures, and wrote the paper and the Supplementary Information. J.P.L. wrote the simulation programs and prepared preliminary versions of the database, some figures and texts.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51335-4.

Correspondence and requests for materials should be addressed to Joaquim Fort.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.