

Supplementary Materials for  
**Past human expansions shaped the spatial pattern of Neanderthal ancestry**

Claudio S. Quilodrán *et al.*

Corresponding author: Claudio S. Quilodrán, [claudio.quilodran@unige.ch](mailto:claudio.quilodran@unige.ch);  
Mathias Currat, [mathias.currat@unige.ch](mailto:mathias.currat@unige.ch)

*Sci. Adv.* **9**, eadg9817 (2023)  
DOI: 10.1126/sciadv.adg9817

**The PDF file includes:**

Tables S1 to S10  
Figs. S1 to S6  
Texts S1 to S4  
Legends for data S1 and S2

**Other Supplementary Material for this manuscript includes the following:**

Data S1 and S2

**Table S1. Best models on the level of Neanderthal ancestry in Eurasia.** The models were selected based on the lowest AIC values. They evaluate the effect of spatial (latitude and longitude) and temporal (date in years BP) variation on the level of Neanderthal ancestry. Groups of populations are represented by hunter gatherers (HGs), Neolithic farmers (FAs), other ancient samples (OTs), and modern samples (MDs). They were collected in different locations in Europe and Asia. A) Full spatiotemporal dataset in Eurasia (~40000 years BP – current time). B) Temporal subset of the full dataset focusing on ancient DNA samples in Europe and Asia. C) Spatial subset of the full dataset focusing on Europe. Neanderthal ancestry was log-transformed (see methods).

Model	Variable	Estimate	SE	<i>t</i>	p-value
A) Full Eurasia ( <i>n</i> = 2625)	Intercept	-4.569	0.085	-53.67	<0.001
	Latitude	0.014	0.002	7.500	<0.001
	Longitude	0.006	0.001	6.119	<0.001
	Time	-1.8E-05	7.3E-06	-2.405	0.016
	Continent:Europe	0.526	0.087	6.057	<0.001
	Latitude x Longitude	-1.1E-04	2.1E-05	-5.086	<0.001
	Latitude x Continent:Europe	-0.011	0.002	-5.930	<0.001
	Latitude x Time	3.0E-07	1.4E-07	2.099	0.036
	Longitude x Continent:Europe	-0.002	0.001	-2.767	0.006
B) Ancient Eurasia ( <i>n</i> = 2534)	Time x Continent:Europe	1.1E-05	3.0E-06	3.745	<0.001
	Intercept	-3.750	0.044	-86.007	<0.001
	Time	-2.6E-05	6.98E-06	-3.671	<0.001
	Population:HG	-0.050	0.0545	-0.921	0.357
	Population:OT	-0.071	0.046	-1.537	0.124
	Continent:Europe	-0.031	0.016	-1.960	0.050
	Time x Population:HG	2.6E-05	7.2E-06	3.560	<0.001
	Time x Population:OT	1.3E-05	8.4E-06	1.534	0.125
C) Europe ( <i>n</i> = 1517)	Time x Continent:Europe	8.0E-06	2.6E-06	3.037	0.002
	Intercept	-3.867	0.055	-69.964	<0.001
	Latitude	-7.9E-04	1.1E-03	-0.724	0.469
	Longitude	-0.02	3.4E-03	-5.731	<0.001
	Time	7.0E-06	2.6E-06	2.631	0.009
	Population:HG	0.122	0.027	4.565	<0.001
	Population:OT	0.023	0.016	1.500	0.134
	Population:MD	0.016	0.053	0.301	0.764
Latitude x Longitude	3.7E-04	6.9E-05	5.406	<0.001	

**Table S2. Average models on the level of Neanderthal ancestry in Eurasia.** The models were averaged based in a cumulative weighted AIC of 90% ( $\sum \omega_i \geq 0.90$ , see Table S4). They evaluate the effect of spatial (latitude and longitude) and temporal (date in years BP) variation on the level of Neanderthal ancestry. Groups of populations are represented by hunter gatherers (HGs), Neolithic farmers (FAs), other ancient samples (OTs), and modern samples (MDs). They were collected in different locations in Europe and Asia. A) Full spatiotemporal dataset in Eurasia (~40000 years BP – current time). B) Temporal subset of the full dataset focusing on ancient DNA samples in Europe and Asia. C) Spatial subset of the full dataset focusing on Europe. Neanderthal ancestry was log-transformed (see methods).

Model	Variable	Estimate	SE	<i>z</i>	p-value
A) Full Eurasia ( <i>n</i> = 2625)	Intercept	-4.613	0.081	56.623	<0.001
	Latitude	0.015	0.002	8.277	<0.001
	Longitude	0.006	0.001	6.223	<0.001
	Time	-2.6E-06	1.7E-06	1.594	0.111
	Continent:Europe	0.518	0.086	6.045	<0.001
	Latitude x Longitude	-1.1E-04	2.1E-05	5.154	<0.001
	Latitude x Continent:Europe	-0.011	0.002	5.819	<0.001
	Latitude x Time	3.0E-13	3.3E-10	0.001	0.999
	Longitude x Continent:Europe	-0.002	5.9E-04	2.939	0.003
	Time x Continent:Europe	1.1E-05	2.9E-06	3.720	<0.001
	Time x Longitude	-4.7E-22	8.5E-15	0.000	1.000
B) Ancient Eurasia ( <i>n</i> = 2534)	Intercept	-3.75	0.044	85.935	<0.001
	Time	-2.6E-05	7.0E-06	3.669	<0.001
	Population:HG	-0.050	0.055	0.921	0.357
	Population:OT	-0.071	0.046	1.537	0.124
	Continent:Europe	-0.031	0.016	1.954	0.051
	Time x Population:HG	2.6E-05	7.2E-06	3.558	<0.001
	Time x Population:OT	1.3E-05	8.4E-06	1.533	0.125
	Continent:Europe x Population:HG	1.1E-04	0.003	0.033	0.974
	Continent:Europe x Population:OT	3.4E-05	0.001	0.024	0.980
	Time x Continent:Europe	8.0E-06	2.6E-06	3.032	0.002
C) Europe ( <i>n</i> = 1517)	Intercept	-3.867	0.056	69.075	<0.001
	Latitude	-8.0E-04	0.001	0.725	0.469
	Longitude	-0.02	3.4E-03	5.726	<0.001
	Time	6.8E-06	2.7E-06	2.537	0.011
	Population:HG	0.125	0.047	2.657	<0.001

Population:OT	0.023	0.020	1.138	0.255
Population:MD	0.017	0.060	0.289	0.772
Latitude x Longitude	3.7E-04	6.9E-05	5.401	<0.001
Latitude x Time	-3.4E-14	1.9E-10	0.000	1.000
Latitude x Population:HG	-6.2E-05	7.4E-04	0084	0.933
Latitude x Population:OT	1.5E-05	2.3E-04	0.063	0.950
Latitude x Population:MD	-3.3E-05	6.0E-04	0.055	0.956
Longitude x Time	-4.0E-14	1.3E-10	0.000	1.000
Longitude x Population:HG	-1.1E-10	7.1E-07	0.000	1.000
Longitude x Population:OT	7.1E-12	2.1E-07	0.000	1.000
Longitude x Population:MD	7.0E-11	8.5E-07	0.000	1.000

---

**Table S3. Full models on the level of Neanderthal ancestry in Eurasia with all tested fixed variable.** They evaluate the effect of spatial (latitude and longitude) and temporal (date in years BP) variation on the level of Neanderthal ancestry. Groups of populations are represented by hunter gatherers (HGs), Neolithic farmers (FAs), other ancient samples (OTs), and modern samples (MDs). They were collected in different locations in Europe and Asia. A) Full spatiotemporal dataset in Eurasia (~40000 years BP – current time). B) Temporal subset of the full dataset focusing on ancient DNA samples in Europe and Asia. C) Spatial subset of the full dataset focusing on Europe. Neanderthal ancestry was log-transformed (see methods).

Model*	Variable	Estimate	SE	<i>t</i>	p-value
A) Full Eurasia ( <i>n</i> = 2625)	Intercept	-4.569	0.085	-53.626	<0.001
	Latitude	0.014	0.002	7.378	<0.001
	Longitude	0.006	0.001	6.122	<0.001
	Time	-1.6E-05	8.4E-06	-1.954	0.051
	Continent:Europe	0.526	0.087	6.060	<0.001
	Latitude x Longitude	-1.1E-04	2.1E-05	-4.942	<0.001
	Latitude x Continent:Europe	-0.011	0.002	-5.823	<0.001
	Latitude x Time	2.9E-07	1.4E-07	2.048	0.041
	Longitude x Continent:Europe	-0.002	5.9E-04	-2.778	0.006
	Longitude x Time	-1.2E-08	4.0E-08	-0.286	0.775
Time x Continent:Europe	1.0E-05	4.0E-06	2.595	0.010	
B) Ancient Eurasia ( <i>n</i> = 2534)	Intercept	-3.738	0.048	-77.140	<0.001
	Time	-2.6E-05	7.2E-06	-3.612	<0.001
	Population:HG	-0.096	0.068	-1.428	0.153
	Population:OT	-0.084	0.050	-1.669	0.095
	Continent:Europe	-0.041	0.030	-1.376	0.169
	Time x Population:HG	2.7E-05	7.3E-06	3.724	<0.001
	Time x Population:OT	1.3E-05	8.4E-06	1.567	0.117
	Continent:Europe x Population:HG	0.052	0.052	1.004	0.316
	Continent:Europe x Population:OT	0.016	0.025	0.620	0.536
Time x Continent:Europe	7.0E-06	3.7E-06	1.950	0.051	
C) Europe ( <i>n</i> = 1517)	Intercept	-3.575	0.177	-20.228	<0.001
	Latitude	-0.007	0.004	-1.818	0.069
	Longitude	-0.021	0.004	-4.878	<0.001

Time	-4.2E-05	2.9E-05	-1.447	0.148
Population:HG	0.608	0.210	2.900	0.004
Population:OT	-0.197	0.131	-1.502	0.133
Population:MD	-0.033	0.306	-0.108	0.914
Latitude x Longitude	3.8E-04	7.5E-05	5.070	<0.001
Latitude x Time	1.0E-06	5.8E-07	1.608	0.108
Latitude x Population:HG	-0.009	0.004	-2.042	0.041
Latitude x Population:OT	0.004	0.003	1.637	0.102
Latitude x Population:MD	-4.9E-04	0.006	-0.077	0.939
Longitude x Time	1.9E-07	3.0E-07	0.638	0.524
Longitude x Population:HG	-0.004	0.002	-1.570	0.117
Longitude x Population:OT	5.2E-04	0.001	0.379	0.704
Longitude x Population:MD	0.005	0.005	1.045	0.296

---

\* Model calls in *nlme* package: A) Full Eurasia : formula = log(f4) ~ (Latitude + Longitude + Time + Continent)^2, random =~ Continent|Population/Period, correlation = corExp(form=~ Latitude + Longitude, nugget=T); B) Ancient Eurasia: formula = log(f4) log(f4)~( Time + Population + Continent)^2, random =~ Continent|Period, correlation = corExp(form=~ Latitude + Longitude, nugget=T); C) Europe: formula = log(f4)~( Latitude + Longitude + Time + Population)^2 - Time:Population, random =~ 1|Period, correlation= corRatio(form=~ Latitude + Longitude, nugget=T). Note that "Period" represents the grouping of samples time within 500-year ranges, and that "Population" exclude modern samples (MD) in the "Ancient Eurasia" model (see Methods).

**Table S4. Candidate models to explain the level of Neanderthal ancestry in Eurasia.** The model selection considered a cumulative weighted AIC of 90% ( $\sum \omega_i \geq 0.90$ ).

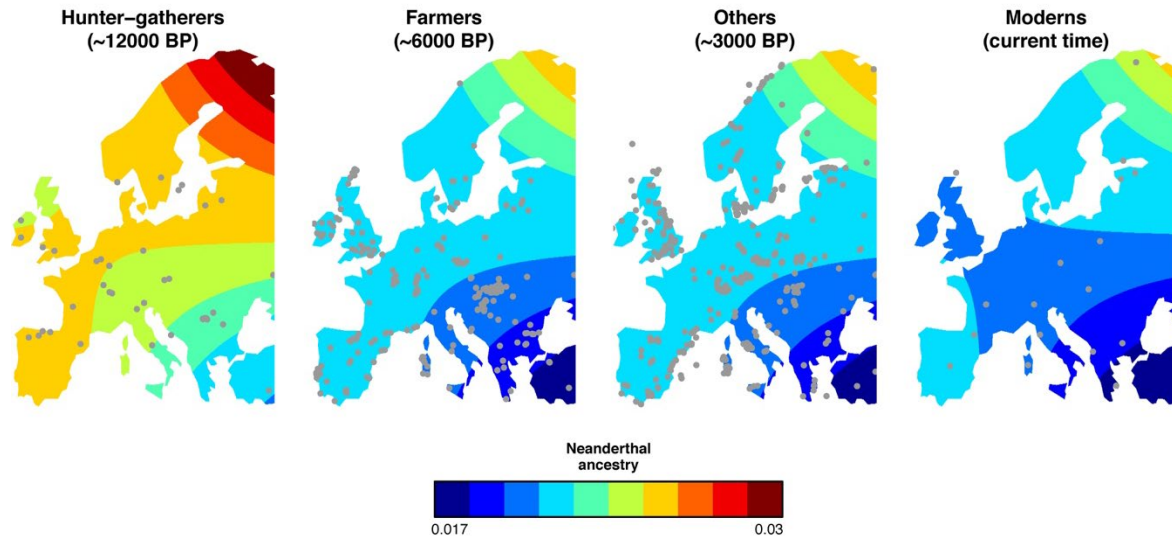
Model title	Models with their variables*	df	AIC	$\Delta$ AIC	$\omega_{AIC}$	$R^2_{GLMM(m)}$	$R^2_{GLMM(c)}$
Full Eurasia	time+lat+continent+long+time:l at+time:continent+lat:continent +lat:long+continent:long	19	-1612.361	0.000	0.484	0.082	0.147
	time+lat+continent+long+time:l at+time:continent+time:long+la t:continent+lat:long+continent:l ong	20	-1610.441	1.920	0.185	0.082	0.147
	time+lat+continent+long+time: continent+lat:continent+lat:long +continent:long	18	-1610.277	2.085	0.171	0.079	0.146
Ancient Eurasia	time+continent+pop+time:conti nent+time:pop	14	-1380.770	0.000	0.663	0.035	0.043
	time+continent+pop+time:conti nent+time:pop+continent:pop	16	-1378.090	2.681	0.173	0.036	0.044
Europe	time+lat+long+pop+lat:long	12	-955.368	0.000	0.162	0.072	0.097
	time+lat+long+pop+time:lat+la t:long+lat:pop	16	-955.236	0.132	0.151	0.077	0.102
	time+lat+long+pop+lat:long+la t:pop	15	-954.884	0.484	0.127	0.076	0.098
	time+lat+long+pop+time:long+ lat:long	13	-953.980	1.388	0.081	0.072	0.097
	time+lat+long+pop+time:lat+ti me:long+lat:long+lat:pop	17	-953.571	1.797	0.066	0.077	0.102
	time+lat+long+pop+time:lat+la t:long	13	-953.509	1.859	0.064	0.072	0.096
	time+lat+long+pop+time:long+ lat:long+lat:pop	16	-953.420	1.948	0.061	0.076	0.098
	time+lat+long+pop+lat:long+lo ng:pop	15	-952.670	2.698	0.042	0.073	0.095
	time+lat+long+pop+time:lat+la t:long+lat:pop+long:pop	19	-952.524	2.844	0.039	0.078	0.100
	time+lat+long+pop+lat:long+la t:pop+long:pop	18	-952.327	3.041	0.035	0.077	0.097
	time+lat+long+pop+time:lat+ti me:long+lat:long	14	-952.146	3.221	0.032	0.072	0.098
lat+long+pop+lat:long+lat:pop	14	-951.687	3.681	0.026	0.074	0.091	

\*Code of each variable: Latitude (lat), Longitude (long), Sample age in YBP (time), Population group (pop), Continental area (continent). The colon represents variables interactions.

**Table S5. Fixed variables considered in the analysis of spatiotemporal variation in Neanderthal ancestry.** Average values and standard deviations (in brackets) are presented. HGs: hunter gatherers, FAs: Neolithic farmers, OTs: other ancient samples, MDs: modern samples. Date is presented in years before present (BP).

	Europe					Asia				
	HG	FA	OT	MD	All	HG	FA	OT	MD	All
Latitude	48.59 (5.65)	47.73 (5.58)	49.47 (6.43)	49.16 (9.17)	48.83 (6.19)	49.97 (9.07)	42.56 (7.55)	44.97 (7.95)	34.19 (13.05)	44.02 (8.84)
Longitude	17.28 (9.66)	9.18 (11.3)	11.46 (8.43)	14.77 (10.83)	11.11 (9.77)	83.95 (43.53)	84.16 (35.68)	75.39 (24.04)	84.69 (31.27)	77.65 (27.66)
Date	10779 (6310)	5911 (1080)	3132 (1342)	0 (0)	4504 (2906)	15752 (11888)	6326 (1544)	2935 (1309)	0 (0)	3657 (3536)





**Figure S1. Spatial variation on the level of Neanderthal ancestry in different population groups across time.** The expected ancestries were computed using the best “Europe” model ( $n = 1517$ ). The grey dots represent the distribution of DNA samples.

## Supplementary Text 1. Theoretical expectations of introgression pattern, investigated with Linear Mixed Models (LMM) and spatially explicit simulations

### *Objective*

The non-homogenous spatial and temporal distribution of paleogenomic samples could raise a question about the power of LMMs to detect spatial introgression gradients emerging following invasive species range expansions, including hybridization with a local species (19). We evaluated this by performing spatially explicit simulations of hybridization during range expansion or hybridization without range expansion, considering either a heterogeneous or a homogeneous distribution of samples throughout space and time.

### *Materials and Methods*

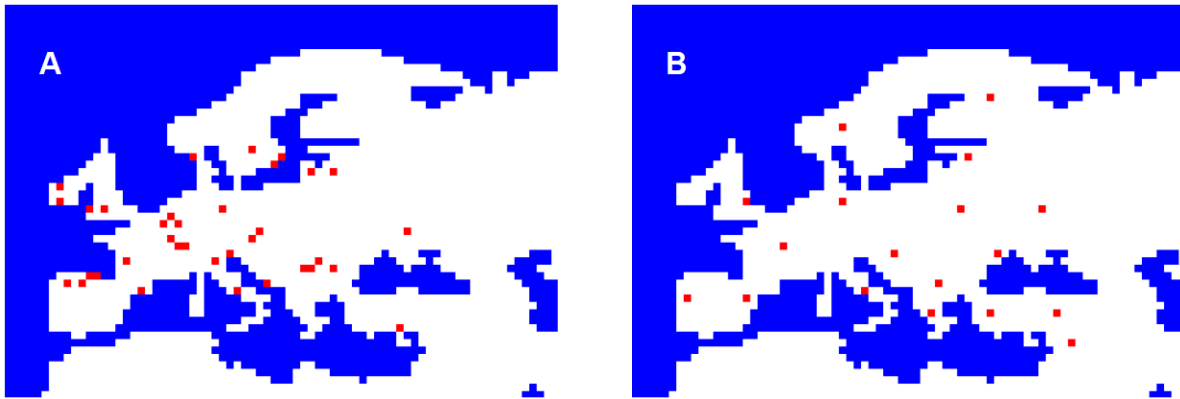
For simulations, we used the software SPLATCHE3 (70), which simulates two interacting species on a map divided into cells of specific dimensions. Simulated individuals can interact only with individuals within the same cell, either of the same or the other species, or can migrate between cells. Each simulated species can have its own demographic characteristics, where different parameters control the gene flow within and between species. The approach consists of two steps: i) a forward-in-time demographic simulation, and ii) a backward-in-time coalescent genetic simulation. The output consists in neutral genetic data for individuals sampled on dates and locations specified by the user.

We performed four sets of 1000 simulations (two sampling procedures and two scenarios, see below), each of which simulating the interaction of two species on a map of Europe. We chose Europe as the simulated area, sampling locations and dates based on hunter-gatherers from the AADR dataset (40). This allowed us to get a rough comparison with the results obtained from the “Europe” model in the main text. Note that simulation parameters were loosely inspired from the expansion of modern humans into Europe and their hybridization with Neanderthals, but we did not try to reproduce this ecological and genetic interaction that would require an independent study on itself. Our aim was to obtain estimates of introgression in a spatio-temporal distribution similar to the empirical dataset to verify the power of LMM to detect spatial introgression gradients, and not to perform parameter inferences.

Simulations used the same framework as (19), but restricting the map to Europe instead of considering the whole Eurasia and simulating a total of 100000 independent SNPs (100 SNPs per simulation) instead of 400000. The dimension of each cell was 100x100 km<sup>2</sup>. All parameter values were the same as in (19):  $K_{loc} = 200$ ,  $K_{inv} = 800$ ,  $m_{loc} = 0.1$ ,  $m_{inv} = 0.2$ , except the hybridization rate ( $\gamma = 0.00375$ ) and the population growth rate ( $r = 0.4$ ). These two parameter values were chosen by performing a set of preliminary simulations to reproduce approximately the level of Neanderthal introgression estimated in modern humans from Europe (~2%) to facilitate interpretation. Simulations last for 2800 generations. The invasive species appears at generation 800 (2000 generations before the present) in the area close to the city of Cairo (chosen arbitrarily), and it takes approximately 300 generations to colonize all the cells on the map. See Data S1 for simulation setting files.

To investigate the effect of the sample’s spatial and temporal distribution on LMM results, we sampled virtual outputs in two different ways: i) *heterogeneous sampling*, where locations and dates of samples were the same as the actual hunter-gatherer samples (HG) from Europe in the main text, and ii) *homogeneous sampling*, where almost (see below) the same number of samples has been distributed uniformly in space and time. We sampled at 35 locations for the heterogeneous dataset. Dates of sampling corresponded to the age of HG samples from the main text, which were

converted to generation number since the start of the simulation by using a generation time of 25 years (Fig. S2). For the homogeneous dataset, we used the same number of samples as in the heterogeneous dataset (plus one so we could sample the same number of times in each location) but spread regularly in space and time within the virtual European map. We thus sampled 18 locations at five different time points, evenly spaced between the earliest and latest observed HG samples. The total number of samples were almost equal for both sampling procedures (89 for the heterogeneous dataset and 90 for the homogenous dataset). Their characteristics (age, coordinates and size) can be found in the files with the extension “.sam” in Data S1.



**Figure S2. Spatial distribution of simulated samples.** A) Heterogeneous dataset, locations identical to Hunter-Gatherer (HG) samples from the main text. B) Homogeneous dataset of the same size as the heterogeneous dataset, plus one (to be able to sample the same number of times in each location).

To investigate whether introgression gradients resulting from hybridization occurring concurrently with range expansion (19) can be detected by LMM, two simulation scenarios were performed: i) *hybridization during the expansion* of the invasive species, while the local species was at equilibrium, and ii) *hybridization without expansion*, with hybridization occurring when both species are at demographic equilibrium. The first scenario considered interspecific competition from the beginning of the expansion (following Currat and Excoffier (7)), with about 20 generations of cohabitation between species before the local one goes extinct. In the second scenario, we let the invasive species colonize the whole area without competition with the local species, until reaching demographic equilibrium (at generation 1180 after the start of the expansion, which corresponds to 100 generations after the end of the expansion). We let hybridization and interspecific competition start at generation 1180 and 1240 respectively, leading to approximately the same amount of introgression in the invasive species as in the first scenario (~2%). Combining the different sampling datasets with simulated scenarios, we get four sets of simulations:

1. Heterogeneous dataset, hybridization during the expansion (1000 stochastic simulations).
2. Heterogeneous dataset, hybridization without expansion (1000 stochastic simulations).
3. Homogeneous dataset, hybridization during the expansion (1000 stochastic simulations).
4. Homogeneous dataset, hybridization without expansion (1000 stochastic simulations).

Although SPLATCHE3 does not calculate  $F_4$ -ratios, it can directly calculate introgression proportions, which is the percentage of loci in one species introgressed from the other (70).  $F_4$ -ratios represent an approximation of introgression proportions, but these are prohibitive to calculate with SPLATCHE3 in the current simulation context, as they require outgroup genomes (i.e., the chimpanzee). Because it has been shown that introgression proportions computed by SPLATCHE3 are highly correlated with values estimated by  $F_4$ -ratios (38), we used the former for the LMM analysis. Average simulated introgression values are presented in Table S6.

**Table S6. Average values of simulated introgression.** These values correspond to scenarios used to explore the power of LMMs to detect spatial gradients of introgression.

	Heterogeneous dataset		Homogeneous dataset	
	Hybridization during expansion	Hybridization without expansion	Hybridization during expansion	Hybridization without expansion
Mean introgression	0.0261	0.0215	0.0252	0.0215
Standard deviation	0.0041	0.0007	0.0038	0.0005

For each simulation, we retrieved the introgression proportions and fitted an LMM. The response variable was the log-transformed introgression proportion. Fixed variables were the latitude and longitude coordinates and their interaction, also including time (age of the sample) as a fixed covariate (in number of generations since the start of the simulation). We included the effect of time period (i.e., grouping dates within 500 years intervals), as a random variable. We also evaluated the spatial and temporal autocorrelation structure for each of the 1000 simulations per scenario in the same way as described in the main text. We used the *nlme* R package (66) to compute the LMMs. We performed model choice of fixed variables with the *MuMIn* R package (68). For simulations with a significant interaction of latitude and longitude, we estimated coefficient trends of these two variables using the *emmeans* R package (67). We plotted simulated introgression levels in Europe with the *raster* R package (69).

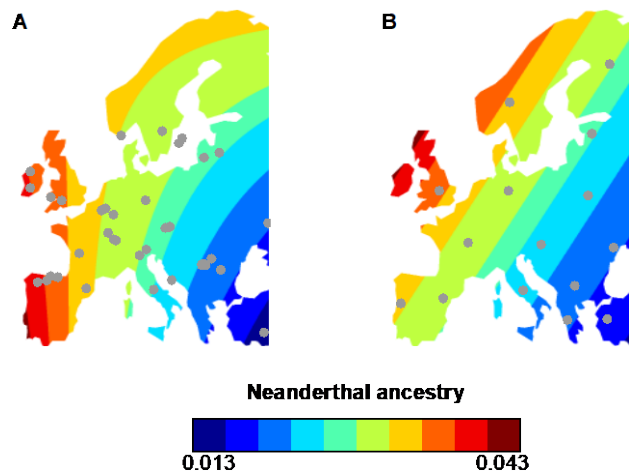
### *Results and Discussion*

We ran LMMs for each of the 1000 simulations and counted how many resulted in significant coefficients for one, two, or none of the two explanatory variables (latitude and longitude). We noted the sign of the significant coefficients (i.e., if introgression levels were positively or negatively related to each explanatory variable). These counts are presented in Table S7. When using a *homogeneous* dataset, the LMM can detect a spatial pattern of introgression in 97.7% of simulations with hybridization during range expansion (positive gradient for latitude and negative gradient for longitude, Table S7). This spatial pattern is detected in only 1.3% of simulations with hybridization without range expansion. When using a *heterogeneous* dataset, similar to the observed one in terms of spatiotemporal distribution, the LMM detects the spatial pattern of introgression in 74.3% of simulations with hybridization during range expansion, but only in 2.2% of simulations with hybridization without range expansion.

**Table S7. Results of simulations exploring the power of LMMs to detect gradients of spatial introgression.** The values correspond to the number of simulations for each scenario that resulted in significant gradients of introgression. In the second column, positive and negative denote the sign of model coefficients at each geographical coordinate. The total number of simulations per scenario is 1000.

Significant variables	Sign of the coefficients	Heterogeneous dataset		Homogeneous dataset	
		Hybridization during expansion	Hybridization without expansion	Hybridization during expansion	Hybridization without expansion
Both variables significant	Both positive	1	7	0	3
	Both negative	4	13	0	1
	Positive longitude, negative latitude	0	16	0	3
	Negative longitude, positive latitude	743	22	977	13
Only longitude significant	Positive	0	79	0	21
	Negative	222	82	15	108
Only latitude significant	Positive	27	100	8	71
	Negative	0	113	0	30
None significant		3	568	0	750

We then averaged introgression proportions among the 1000 simulations and used these mean values to fit two LMMs, one for each of the two set of simulations with *hybridization during the expansion*. Figure S3 shows the expected introgression proportions across Europe, confirming that LMMs can project a spatial pattern of introgression on simulated data like the one observed with the empirical dataset in the main text (first panel of Fig. S1).



**Figure S3. Spatial variation of simulated introgression levels.** Expected values were computed using LMMs fitted with simulated introgression proportions (averaged over 1000 simulations) obtained with the scenario *hybridization during the expansion*. Grey dots represent the distribution of simulated samples. A) Introgression levels from simulations using a heterogeneous dataset. B) Introgression levels from simulations using a homogeneous dataset.

### *Conclusion*

Our simulations demonstrate that LMMs can reliably detect a spatial gradient of introgression arising following range expansion with hybridization. While a homogenous spatiotemporal distribution of samples improves the detection of spatial gradients of introgression, LMMs could also recover significant signals when the sampling is heterogenous, as the one used in the main text. Note that the simulated dataset is more limited in size than the empirical one used in the main text since only hunter-gatherer (HG) samples were included in the simulation. A larger sample size could improve the power of this approach. When simulating hybridization during range expansion, more than 70% of simulations resulted in significant spatial gradients of introgression that are consistent with the empirical observations. This value decreased to less than 2.5% when hybridization is not occurring during a range expansion.

## Supplementary Text 2. Effect of topographic distance from a putative OOA origin in East Africa on the level of Neanderthal ancestry

### *Objective*

In this supplementary text, the aim was to see whether results from the “Full Eurasia” model in the main text could be confirmed by taking distance from Africa as the explanatory variable, instead of latitude and longitude.

### *Materials and Methods*

We analysed the effect of the distance from a putative source of the Out-of-Africa (OOA) range expansion on the level of Neanderthal ancestry in Europe and Asia. We performed a similar analysis to the “Full Eurasia” model in the main text but replacing latitude and longitude by the topographic distance between each sample and a putative OOA origin in East Africa, taking Addis-Abeda as an arbitrary reference (9.03N, 38.74E), following Sanchez-Mazas *et al.* (71). We called this model “Full Eurasia distance”. The topographic distance was computed in meters using the *TopoDistance* R package (72), which computes the shorter topographic distance between pairwise locations by considering a digital elevation map (DEM). This elevation map was obtained with the *elevatr* R package (73).

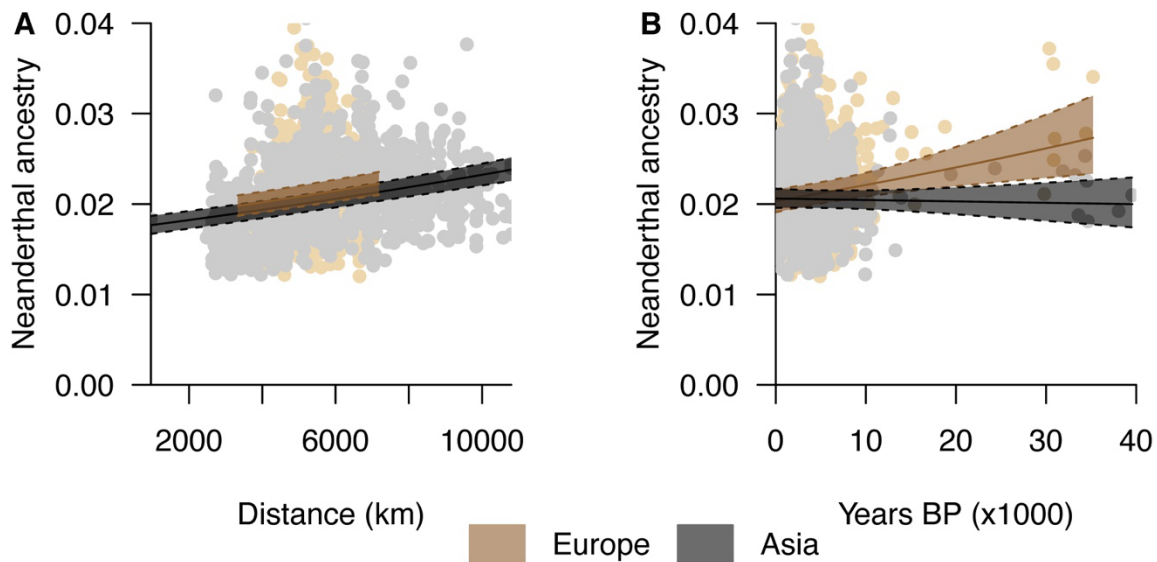
The response variable was the log-transformed  $F_4$ -ratio. Explanatory variables were the topographic distance, time (years BP), continental area (Europe or Asia), and their interactions. We evaluated the effect of population groups (HG: hunter-gatherers, FA: farmers, OT: others, and MD: moderns), as well as the nested effect of time periods within these population groups (i.e., grouping dates by 500 years intervals) as random variables. We also evaluated the influence of the spatial and temporal autocorrelation within the dataset. The selection of the random and fixed structures was performed as described in the main text. The final LMM considered the period nested in population groups as a random intercept and slope of the continental area, as well as an exponential spatial autocorrelation.

### *Results and Discussion*

A pattern similar to the one presented for the “Full Eurasia” model in the main text was obtained when replacing latitude and longitude with the distance from a putative OOA origin in East Africa. The level of Neanderthal ancestry increased with the distance from Africa, as expected after a range expansion with hybridization (19) (Fig. S4). The interaction between time (years BP) and continental area (Europe and Asia) was also significant (Table S8). While the level of Neanderthal ancestry was higher in Europe when compared to Asia during the Palaeolithic, this level decreased with time much faster in the former region, resulting in a slightly higher level in Asia after the Neolithic transition (Fig. S4). While this approach simplified our model by reducing spatial coordinates to a single variable, it implies the definition of a geographic origin for the OOA, which is still difficult to determine (43, 74-76).

**Table S8. Best LMM on the level of Neanderthal ancestry in Eurasia considering the distance from a putative location in East Africa.** The model was selected based on the lowest AIC values. It evaluates the effect of spatial (topographic distance in meters) and temporal (date in years BP) variations on the log-transformed level of Neanderthal ancestry. This LMM is called “Full Eurasia distance” ( $n = 2625$ ).

Variable	Estimate	SE	$t$	p-value
Intercept	-4.062	0.032	-126.702	<0.001
Distance	3.0E-08	2.9E-09	10.457	<0.001
Time	-1.0E-06	1.9E-06	-0.405	0.686
Location:Europe	0.004	0.023	0.176	0.861
Time x Location:Europe	9.2E-06	2.8E-06	3.228	0.001



**Figure S4. Effects of the topographic distance from a putative OOA origin in East Africa and time on the level of Neanderthal ancestry in both Europe and Asia.** A) Effect of topographic distance, and B) effect of time. The solid and dotted lines represent the estimated values and 95% confidence intervals. The coloured dots represent the distribution of the full dataset of ancient and modern DNA samples used in the “Full Eurasia distance” analysis ( $n = 2625$ ).



### Supplementary Text 3. Neanderthal ancestries estimated from a subset of data less affected by selection

#### *Objective*

The main aim was to check the robustness of our results with a subset of data assumed to be less subject to selection.

#### *Materials and Methods*

We computed Neanderthal ancestries on genomes from the AADR dataset using a subset of genomic regions less affected by natural selection. We followed Pouyet *et al.* (63), who proposed two lists of genomic regions particularly suited for demographic inference. The first list considers regions with recombination rates above 1.5 cM/Mb and mutations that are GC-conservative. The second list is similar to the first one, but removes regions that are located less than 100bp from conserved elements. Both lists reduced the mean number of SNPs available for computing  $F_4$ -ratios by about 99.8%, so we were not able to compute spatiotemporally distributed  $F_4$ -ratios with these lists of neutral genomic regions (only 13 and 4 genomes remained from the full Eurasian database). By relaxing the condition of mutations that are GC-conservative in the second list, we were able to estimate significant  $F_4$ -ratios in about 50% of paleogenomic samples, while using only about 20% of available SNPs. According to Pouyet *et al.* (63), because of the higher recombination rate and distance to conserved elements, this dataset can be considered to be less affected by background selection, but is still affected by GC-biased gene conversion

We computed  $F_4$ -ratios in the same way as described in the methods in the main text. This variable was included as the response variable in a model similar to the “Full Eurasia” model in the main text. We called this model “Full Eurasia neutral”. To maintain the Gaussian distribution of residuals, the response variable was logit-transformed after rescaling the values to levels close to zero (0.001) and close to 1 (0.999). We used the *scale* R package for rescaling the variable and the *car* R package for the logit transformation. Explanatory variables were the latitude, longitude, time (years BP), continental area (Europe or Asia), and their interactions. We evaluated the effect of population groups (HG: hunter-gatherers, FA: farmers, OT: others, and MD: moderns), as well as the nested effect of the time period within these population groups (i.e., grouping dates by 500 years intervals) as random variables. We also evaluated the influence of the spatial and temporal autocorrelation within the dataset. The random and fixed structures were selected in the same way as described in the main text. The final LMM considered population groups as a random intercept, as well as a spherical spatial autocorrelation structure.

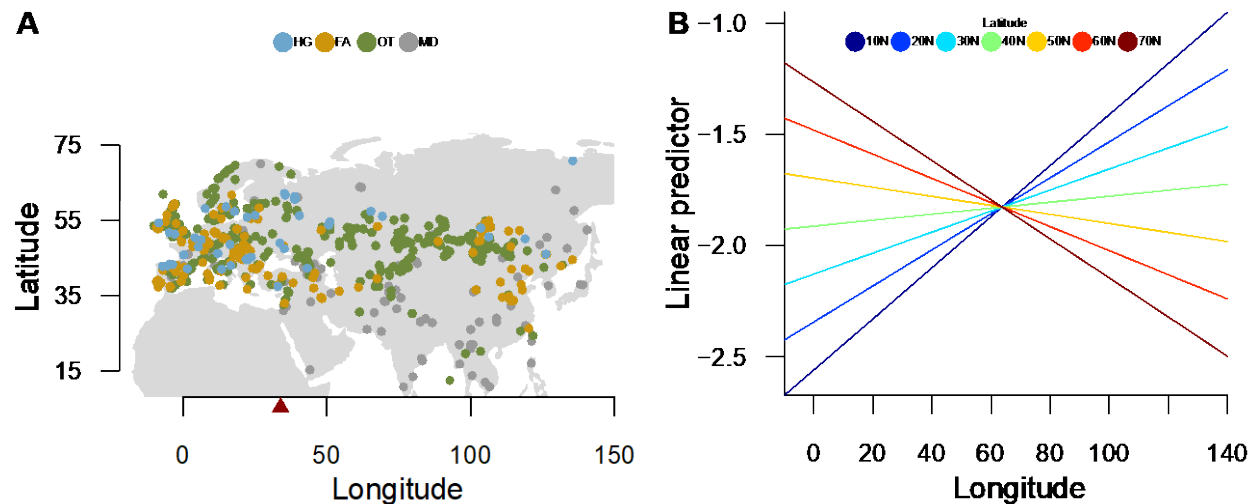
#### *Results and Discussion*

A similar pattern to the one found for the “Full Eurasia” model was obtained for the “Full Eurasia neutral” model. The interaction of latitude and longitude was significant, meaning that the Neanderthal ancestry changed with both variables, as well as the interaction between continental area and latitude (Table S9). Time is no longer significant, as well as the interaction between longitude and continent. However, the slopes of latitude and longitude changed with the values of each other variable. While southern latitudes, which are represented by more samples from Asia, tend to have a positive slope for longitude, northern latitudes, which are represented by more samples from Europe, tend to have a negative slope for longitude (Fig. S5). While we cannot exclude that background selection might have influenced some of the results observed with the full dataset, using more neutral regions only allows us to keep around 10% of the original dataset,

implying a large loss of power for analysis. For this reason, we favoured the use of the full dataset in the main text.

**Table S9. Best LMM explaining the level of Neanderthal ancestry in Eurasia by considering a subset of data less affected by selection.** The model was selected based on the lowest AIC values. The computation of  $F_4$ -ratios is restricted to genomic regions with a recombination rate higher than 1.5 cM/Mb and at least 100bp distance from conserved elements. The level of Neanderthal ancestry was logit-transformed ( $n = 1190$ ).

Variable	Estimate	SE	$t$	p-value
Intercept	-3.328	0.486	-6.841	<0.001
Latitude	0.033	0.009	3.678	<0.001
Longitude	0.015	0.005	3.246	0.001
Location:Europe	1.106	0.420	2.636	0.008
Latitude x Longitude	-3.4E-04	1.0E-04	-3.366	<0.001
Latitude x Location:Europe	-0.024	0.009	-2.665	0.008



**Figure S5. Expected ancestry estimated with a more neutral subset of the data. A)** Distribution of samples in Eurasia. Coloured dots represent palaeogenomic samples of hunter-gatherers (HG,  $n = 83$ ), early farmers (FA,  $n = 250$ ), other ancient (OT,  $n = 771$ ), and modern (MD,  $n = 86$ ) samples. The red triangle represents the longitudinal limit ( $34^\circ$ ), which in our study separates European ( $n=636$ ) from Asian ( $n=554$ ) population samples. **B)** Longitudinal trends with various latitudinal values in our best “Full Eurasia neutral” model. The response variable is the logit transformed  $F_4$ -ratio ( $n = 1190$ ).

## Supplementary Text 4. Impact of Neolithic Farmer ancestry on the level of Neanderthal ancestry in Europe

### *Objective*

The aim was to confirm the observation made with the “Europe” model investigated in the main text, according to which Neanderthal ancestry decreased with the arrival of Neolithic farmers in Europe. We assessed in European hunter-gatherer and farmer samples the relationship between the proportion of Neanderthal ancestry and the proportion of early farmer ancestry from Anatolia, thought to be the source of expanding farming populations towards Europe (e.g., 32, 33, 35, 47).

### *Materials and Methods*

We focused this analysis on European paleogenomes. For the 715 European paleogenomes classified either as farmers (FA) or as hunter-gatherers (HG) and showing significant Neanderthal introgression (estimated using  $F_4$ -ratio), we estimated their ancestry from Anatolian farmers (AFA). We used the *qpadm* function of the software *ADMIXTOOLS 2* (61). We followed the procedure described in Lazaridis *et al.* (77), where each genome is modelled as a 1-way, 2-way or 3-way admixture between all possible combinations of three source populations: 1) Anatolian Neolithic farmers (AFA), 2) Western hunter-gatherers (WHG), and 3) steppe Bronze Age (SBA). We retained for each genome the model with the highest  $P$ -value. This analysis required the definition of various genomes considered either as source or outgroup populations for each source of admixture. Genomes included in one of these groups were excluded from that group when estimating their own AFA ancestry, as suggested by Lazaridis *et al.* (77). The group identity (source, outgroup, neither) of each sample is presented in Data S2, which also presents the number of retained source populations (1-way, 2-way, or 3-way) for estimating the AFA ancestry. When the chosen model did not include AFA ancestry, we set the AFA value to zero. We excluded genomes that could not be modelled as an admixture between these three source populations (i.e., none of the three models giving a significant  $P$ -value) from the analysis. We averaged the estimated AFA proportion for samples with the same date, geographic coordinates, and population group. This resulted in a final dataset of 214 samples, ranging from about 3800 to 10200 years BP.

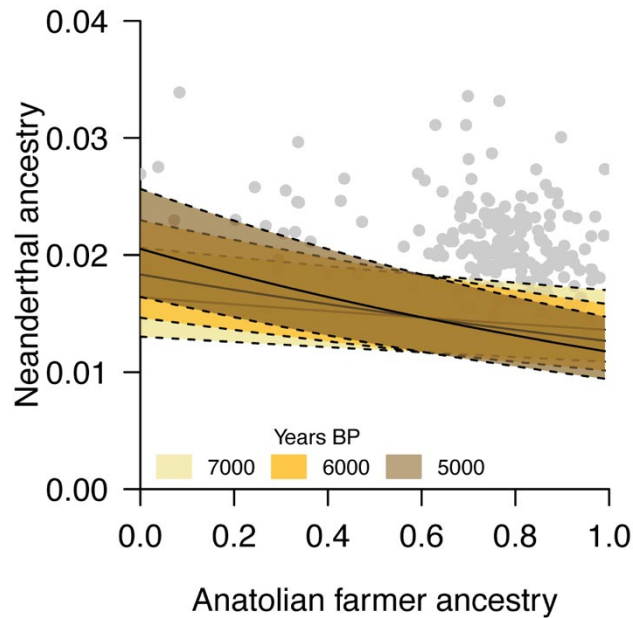
We developed an LMM by following the same procedure as described in the main text. The response variable was the log-transformed  $F_4$ -ratio. Fixed variables were the AFA ancestry, time (years BP), and their interaction. We evaluated the random effect of the time period (i.e., grouping dates within 500-year intervals) and the influence of the spatial and temporal autocorrelation within the dataset. Because the random effect of the time period was not retained after model selection with AIC values, the final model was a GLS (generalized least square) that considered a Gaussian spatial autocorrelation.

### *Results and Discussion*

The Anatolian farmer ancestry in European paleogenomes is negatively related to the Neanderthal ancestry (Table S10). Because the interaction between AFA ancestries and the date of samples is significant, the negative slope between AFA and Neanderthal ancestries also changes with time (Table S10). This means the strength of this negative relationship increases with time following the appearance and expansion of Neolithic farmers. (Fig. S6).

**Table S10. Best GLS explaining the impact of Anatolian farmer ancestry on the level of Neanderthal ancestry in Europe.** The model was selected based on the lowest AIC values. It evaluates the effect of Anatolian farmer ancestry, time (date in years BP) of the sample, and their interaction on the log-transformed level of Neanderthal ancestry. The dataset is focused on farmers (FA) and hunter-gatherers (HG) in Europe ( $n = 214$ ).

Variable	Estimate	SE	<i>t</i>	p-value
Intercept	-3.323	0.127	-26.236	<0.001
Anatolian farmer ancestry	-1.490	0.057	-26.195	<0.001
Time	-1.1E-04	1.2E-05	-9.273	<0.001
Time x Anatolian farmer ancestry	1.9E-04	1.2E-05	14.916	<0.001



**Figure S6. Impact of the Anatolian farmer ancestry on the level of Neanderthal ancestry in European paleogenomes.** Solid and dotted lines represent the estimated values and 95% confidence intervals, respectively. Different colours represent the slope of the effect of Anatolian farmer ancestry at three different time periods. Dots represent the distribution of farmers (FA) and hunter-gatherers (HG) in Europe ( $n = 214$ ) present within the dataset.

**Data S1. Computer simulations data files.** Setting files and executable to conducting spatially explicit simulations with SPLATCHE3.

**Data S2. Genomes included in the present study.** Population group: group associated to each sample depending on its cultural background and age (HG: hunter gatherers, FA: Neolithic farmers, OT: other ancient samples, MD: modern samples). Continental region: if the sample was located in Asia or Europe. Master IDs: the Master ID in the AADR database. Index: the indices in the information files of the AADR database. Source and outgroup: genome that were used as an outgroup or source population for computing Anatolian farmer proportions. Publications: the publications of the studies that produced the genomes.