S1 Supplementary Methods

S1.1 Expected genetic dissimilarities in population genetics

Here we derive the expected genetic dissimilarity $\mathrm{E}\{D_{ij}\}$ between two distinct individuals i and j, as a function of the expected coalescence times between the alleles carried by each individual. In population genetics, the expected coalescence time can be considered a distance metric: the larger the coalescent time is, the more differentiated i and j are expected to be as they would not share mutations either lineage accumulates after the split from their most recent common ancestor (MRCA). We start with the simplest case: haploid SNPs, and then we extend to diploid SNPs and microsatellites.

S1.1.1 Haploids

Following [McVean, 2009], we condition on the event S that the SNP segregates in a sample, and we take the limit $\theta \to 0$, where θ is the mutation rate. If the samples are haploid, the event S is equivalent to observing exactly one mutation. Let $Z_i \in \{0, 1\}$ denote the allele carried by individual i at a random SNP. The expected value of the genetic difference $(Z_i - Z_j)^2$ is the probability that $Z_i \neq Z_j$, and so

$$\mathrm{E}\Big\{(Z_i - Z_j)^2 \,|\, S\Big\} = \Pr\Big\{Z_i \neq Z_j \,|\, S\Big\} = \Pr\{Z_i \neq Z_j\} / \Pr\{S\} \tag{S1a}$$

$$= \lim_{\theta \to 0} \frac{\mathrm{E}\Big\{(\theta t_{ij}) \exp(-\theta t_{ij})\Big\}}{\mathrm{E}\Big\{(\theta t_{\mathrm{tot}}/2) \exp(-\theta t_{\mathrm{tot}}/2)\Big\}} = \frac{2T_{ij}}{T_{\mathrm{tot}}}, \tag{S1b}$$

where the t's are (random) coalescence times and the T's are their expectations. For a given sample, T_{tot} denotes the expected total length (sum of all branches) of a random genealogy relating the individuals to the sample's MRCA and T_{ij} the time to the recent common ancestor of the alleles carried by individual i and j. Intuitively, $\Pr\{Z_i \neq Z_j\}$ is proportional to $2T_{ij}$ because i and j carry different alleles if and only if a mutation occurs on the path from i to j through the pair's MRCA, which is of expected length $2T_{ij}$.

S1.1.2 Diploids

We model the genotype of a diploid individual as the sum of two haploids: $Z_i = Z_{i_1} + Z_{i_2} \in \{0, 1, 2\}$, where the subscript indicates one of two alleles. To derive the expected genetic dissimilarity between two distinct individuals i and j, we use the results from [McVean, 2009] that (conditioning on the event S and taking the limit $\theta \to 0$ as in the haploid case):

$$\mathrm{E}\Big\{Z_{i_1}^2 \,|\, S\Big\} = \mathrm{E}\Big\{Z_{i_1} \,|\, S\Big\} = T_{\mathrm{mrca}}/T_{\mathrm{tot}}, \tag{S2a}$$

$$\mathrm{E}\Big\{Z_{i_1} Z_{j_1} \,|\, S\Big\} = \Big(T_{\mathrm{mrca}} - T_{ij}\Big)/T_{\mathrm{tot}}, \tag{S2b}$$

where T_{mrca} denotes the expected coalescence time to the MRCA of all sampled individuals, i.e., the height of the average genealogy of the sample. Thus, we have:

$$\mathrm{E}\Big\{(Z_i - Z_j)^2 \,|\, S\Big\} = \mathrm{E}\Big\{Z_{i_1}^2 + Z_{i_2}^2 + Z_{j_1}^2 + Z_{j_2}^2 + 2Z_{i_1}Z_{i_2} + 2Z_{j_1}Z_{j_2} - 2Z_{i_1}Z_{j_1} - 2Z_{i_1}Z_{j_2} - 2Z_{i_2}Z_{j_1} - 2Z_{i_2}Z_{j_2} \,|\, S\Big\} \tag{S3a}$$

$$= 4\frac{T_{\mathrm{mrca}}}{T_{\mathrm{tot}}} + 2\Big(\frac{T_{\mathrm{mrca}} - T_i}{T_{\mathrm{tot}}}\Big) + 2\Big(\frac{T_{\mathrm{mrca}} - T_j}{T_{\mathrm{tot}}}\Big) - 8\Big(\frac{T_{\mathrm{mrca}} - T_{ij}}{T_{\mathrm{tot}}}\Big) = 2\Big(\frac{4T_{ij} - T_i - T_j}{T_{\mathrm{tot}}}\Big), \tag{S3b}$$

where $T_i = \mathrm{E}\{t_{i_1 i_2}\}$ and $T_j = \mathrm{E}\{t_{j_1 j_2}\}$.

S1.1.3 Microsatellites

At a microsatellite locus, an allele is coded as the "number of repeats" of a short DNA motif (two to six base pairs). The mutation process at this locus can be modeled by a symmetric stepwise mechanism where the number of repeats increases or decreases by 1, with equal probability [Ohta and Kimura, 1973].

We model the genotype Z_{il} of a diploid individual at microsatellite locus l as the average of the two alleles, Z_{i_1} and Z_{i_2}. Under the symmetric mutation process,

$$Z_{i_1} = a_l + \sum_{k=1}^{K_{i_1}} S_k, \tag{S4}$$

where a_l is the ancestral allele at microsatellite locus l (the allele carried by the MRCA of all lineages in the sample), K_{i_1} denotes the number of mutations that occur on the lineage from i_1 to the MRCA, and the S_ks are independent binary random variables with $\Pr\{S_k = 1\} = \Pr\{S_k = -1\} = \frac{1}{2}$.

Let θ_l denote the mutation rate at locus l. If we assume that the mutations S_k at locus l occur as a Poisson process with mutation rate θ_l, then $K_{i_1} \mid t_{ij} \sim \mathrm{Po}(\theta_l t_{\mathrm{mrca}})$ [Hudson, 1990]. Thus, we have:

$$\begin{aligned}
\mathrm{E}\Big\{Z_{i_1}^2 \mid \theta_l\Big\} &= a_l^2 + \mathrm{E}\Big\{\mathrm{E}\Big\{\Big(\sum_{k=1}^{K_{i_1}} S_k\Big)^2 \mid t_{\mathrm{mrca}}\Big\} \mid \theta_l\Big\} = a_l^2 + \mathrm{E}\Big\{\mathrm{E}\Big\{\sum_{k=1}^{K_{i_1}} S_k^2 \mid t_{\mathrm{mrca}}\Big\} \mid \theta_l\Big\} \\
&= a_l^2 + \mathrm{E}\Big\{\mathrm{E}\Big\{K_{i_1} \mid t_{\mathrm{mrca}}\Big\} \mid \theta_l\Big\} = a_l^2 + \mathrm{E}\Big\{\theta_l t_{\mathrm{mrca}} \mid \theta_l\Big\} = a_l^2 + \theta_l T_{\mathrm{mrca}}.
\end{aligned} \tag{S5}$$

(We have used: $\mathrm{E}\{Z_{i_1}\} = a_l + \sum_k S_k = a_l$ and $\mathrm{E}\{S_k S_{k'}\} = 0$ since the S_ks have mean 0, variance 1 and are independent.) Similarly,

$$\mathrm{E}\Big\{Z_{i_1} Z_{j_1} \mid \theta_l\Big\} = a_l^2 + \theta_l\Big(T_{\mathrm{mrca}} - T_{ij}\Big). \tag{S6}$$

Now we can use (S5) and (S6) to obtain:

$$\mathrm{E}\Big\{(Z_i - Z_j)^2 \mid \theta_l\Big\} = \frac{\theta_l}{2}\Big(4T_{ij} - T_i - T_j\Big). \tag{S7}$$

Compared to (S3), there is a factor of 1/4 because we model a diploid genotype at a microsatellite locus as the *average* of two alleles rather than the *sum* of two alleles, as in the case of SNPs.

S1.2 Expected genetic dissimilarities in EEMS

EEMS is based on the stepping stone model [Kimura and Weiss, 1964], which specifies that

$$T_{ij} = T_{\delta(i)\delta(j)} \tag{S8}$$

for individuals i and j drawn randomly from demes $\delta(i)$ and $\delta(j)$, respectively.

Let D be the matrix of observed genetic differences: $D = (D_{ij}) = \Big((Z_i - Z_j)^2\Big)$. If $i = j$, both the observed and the expected dissimilarity with self is 0. If $i \neq j$, the expected genetic dissimilarity is given by equations (S1), (S3) and (S7):

$$\mathrm{E}\{D_{ij} \mid *\} = \begin{cases} \sigma^2 T_{\delta(i)\delta(j)}, \text{ hap/SNP}; & \sigma^2\Big(4T_{\delta(i)\delta(j)} - T_{\delta(i)} - T_{\delta(j)}\Big), \text{ dip/SNP}; \\ \sigma_l^2 T_{\delta(i)\delta(j)}, \text{ hap/sat } l; & \sigma_l^2\Big(4T_{\delta(i)\delta(j)} - T_{\delta(i)} - T_{\delta(j)}\Big), \text{ dip/sat } l; \end{cases} \tag{S9}$$

where the symbol $*$ indicates the event that the site segregates in the sample (if the markers are SNPs), and the mutation rate θ_l (if the locus is microsatellite l). The constant of proportionality is $2/T_{\text{tot}}$ for SNPs and $\theta_l/2$ for microsatellites, which we write σ^2 and σ_l^2, respectively, because the size of the average sample genealogy T_{tot} and the mutation rate θ_l are not of interest here.

Under the stepping stone model, the expected coalescence time between two distinct demes α and β can be approximated by splitting $T_{\alpha\beta}$ into two components:

$$T_{\alpha\beta} - \big(T_\alpha + T_\beta\big)/2 \approx R_{\alpha\beta}/4 \tag{S10a}$$

$$\big(T_\alpha + T_\beta\big)/2 \approx \big(q_\alpha + q_\beta\big)/2 \tag{S10b}$$

where $R_{\alpha\beta}$ is the resistance distance between α and β, which depends on the effective migration rates m, and q_α, q_β are the effective diversity rates. In this approximation, resistance distances R specify the expected genetic differentiation between distinct demes in the habitat (the between-demes component), while the effective diversity rates q specify the expected genetic differentiation between distinct individuals from the same deme (the within-deme component).

EEMS uses the approximation given by (S10) to specify a model for the expected genetic dissimilarities:

$$\mathrm{E}\{D_{ij} \,|\, *\} = \begin{cases} \sigma^2\Big(R_{\delta(i)\delta(j)}/4 + (q_{\delta(i)} + q_{\delta(j)})/2\Big), \text{ hap/SNP}; & \sigma^2\Big(R_{\delta(i)\delta(j)} + (q_{\delta(i)} + q_{\delta(j)})\Big), \text{ dip/SNP}; \\ \sigma_l^2\Big(R_{\delta(i)\delta(j)}/4 + (q_{\delta(i)} + q_{\delta(j)})/2\Big), \text{ hap/sat } l; & \sigma_l^2\Big(R_{\delta(i)\delta(j)} + (q_{\delta(i)} + q_{\delta(j)})\Big), \text{ dip/sat } l; \end{cases} \tag{S11}$$

Therefore, $\mathrm{E}\{D \,|\, *\} \propto \Delta(m, q)$ and the constant of proportionality is σ^2 for SNPs and σ_l^2 for microsatellite l. Furthermore, the expected dissimilarity matrix $\Delta(m, q)$ has the same form in all four cases, and for the purpose of generality, we can write

$$\Delta(m, q) = JBJ' + \tfrac{1}{2}Jw1_n' + \tfrac{1}{2}1_n w'J' - W_n, \tag{S12}$$

where B is the matrix of between-deme dissimilarities, w is the vector of within-deme dissimilarities, $J = (J_{i\alpha})$ is an indicator matrix such that $J_{i\alpha} = 1$ if individual i comes from deme α, and $W_n = \mathrm{diag}\{Jw\}$. [We subtract the diagonal matrix $W_n = \mathrm{diag}\{\frac{1}{2}Jw1_n' + \frac{1}{2}1_n w'J'\}$ because the dissimilarity matrix $\Delta(m, q)$ has a main diagonal of 0s.] If there are n individuals sampled from o demes, then $J \in \mathbb{Z}^{n\times o}$, $B \in \mathbb{R}^{o\times o}$, $w \in \mathbb{R}^o$. To simplify the notation, we drop the subscripts and write plainly 1 for the vector of 1s. The dimension will be clear from the context because $B = (B_{\alpha\beta})$ is an $o \times o$ matrix and $\Delta(m, q)$ is an $n \times n$ matrix.

S1.3 Computing the Wishart log likelihood $\ell(k, m, q, \sigma^2)$

EEMS represents the population as a connected undirected graph (V, E), with effective migration rates $m = \{(\alpha, \beta) \in E : m_{\alpha\beta}\}$ and effective diversity rates $q = \{\alpha \in V : q_\alpha\}$. Furthermore, EEMS models the observed genetic differences D between n individuals, averaged across p SNPs, through a positive definite transformation:

$$-LDL' \,|\, k, m, q, \sigma^2 \sim \mathrm{W}_{n-1}\Big(k, -\frac{\sigma^2}{k}L\Delta(m, q)L'\Big), \tag{S13}$$

where $\Delta(m, q)$ is the expected genetic dissimilarity matrix, given by equation (S12), L is a $(n-1) \times n$ basis for contrasts on n elements, k is the degrees of freedom, constrained to lie in the range $[n, p]$, and σ^2 is a scale parameter. (See Section S1.2 for a demographic interpretation of σ^2.)

By definition, a contrast is a linear combination with coefficients that add to zero, so $L1 = 0$ and

$$L\Delta L' = L\Big(JBJ' + \tfrac{1}{2}Jw1' + \tfrac{1}{2}1w'J' - W_n\Big)L' = L\Big(JBJ' - W_n\Big)L. \tag{S14}$$

This implies that Δ and $JBJ' - W_n$ are equivalent under the Wishart likelihood (S13) because they give the same likelihood. Therefore, without loss of generality, we can assume that the expected dissimilarity matrix has the form:

$$\Delta = JBJ' - W_n, \tag{S15}$$

where JBJ' is a block matrix and W_n is a diagonal matrix. We can exploit this structure to compute the Wishart log likelihood efficiently, without explicitly constructing the $n \times n$ matrix Δ. Thus the computational cost scales with the grid size, not with the number of samples. The hard-to-compute terms of the log likelihood (S13) are the determinant and the trace:

$$\operatorname{tr}\Big\{(L\Delta L')^{-1}LDL'\Big\} = \operatorname{tr}\Big\{\Delta^{-1}\Big(\Delta L'(L\Delta L')^{-1}L\Big)D\Big\} = \operatorname{tr}\Big\{\Delta^{-1}QD\Big\}, \tag{S16a}$$

$$\det\Big\{-(L\Delta L')^{-1}\Big\} = \det\Big\{-L'(L\Delta L')^{-1}L\Big\}/\det\Big\{LL'\Big\} = \operatorname{Det}\Big\{-\Delta^{-1}Q\Big\}/\det\Big\{LL'\Big\}, \tag{S16b}$$

where det denotes the standard determinant (the product of all eigenvalues), Det denotes the pseudo determinant (the product of the nonzero eigenvalues) and

$$Q = \Delta L'(L\Delta L')^{-1}L = I - 1(1'\Delta^{-1}1)^{-1}1'\Delta^{-1} \tag{S17}$$

is an orthogonal projection matrix with kernel $\{1\}$, the space of constant functions.

The distance matrix $\Delta = JBJ' - W_n$ is the sum of a block matrix and a diagonal matrix, and its inverse Δ^{-1} has similar "almost-block" structure:

$$\Delta^{-1} = JXJ' - W_n^{-1}, \tag{S18}$$

where X is an unknown $o \times o$ matrix. Since $\Delta\Delta^{-1} = I$, the solution X must satisfy:

$$JBCXJ' - W_nJXJ' - JBJ'W_n^{-1} + W_nW_n^{-1} = I \qquad \Leftrightarrow \qquad J(BC - W_o)XJ' = JBW_o^{-1}J', \tag{S19}$$

where $C = J'J = \operatorname{diag}\{n_\alpha\}$ is the diagonal matrix of sample counts n_α. Since every term in equation (S19) has exact block structure which depends on the sample configuration through J, it is sufficient to solve the lower-dimensional problem:

$$\Big(BC - W_o\Big)X = BW_o^{-1}. \tag{S20}$$

This is a system of linear equations for the unknown X as a function of the between-demes dissimilarities B, the within-deme dissimilarities $W_o = \operatorname{diag}\{w\}$ and the sample counts C. The unknown X can be computed efficiently without a matrix inversion, by performing the LU factorization of $Y = BC - W_o$.

We can express the pseudo-determinant $\operatorname{Det}\{-\Delta^{-1}Q\}$ and the trace $\operatorname{tr}\{\Delta^{-1}QD\}$ in terms of the auxiliary matrix X. Using the definition of the orthogonal projection Q in equation (S17) and the properties of the trace,

$$\operatorname{tr}\Big\{\Delta^{-1}QD\Big\} = \operatorname{tr}\Big\{\Delta^{-1}D\Big\} - \frac{1}{1'\Delta^{-1}1}\operatorname{tr}\Big\{11'\Delta^{-1}D\Delta^{-1}\Big\}. \tag{S21}$$

We consider each of these terms in turn. For simplicity of notation, let $c = \mathrm{diag}\{C\}$ and $v = (w_\alpha^{-1})$. Then

$$1'\Delta^{-1}1 = 1'\Big(JXJ' - W_n^{-1}\Big)1 = c'\Big(Xc - v\Big), \tag{S22a}$$

$$\mathrm{tr}\Big\{\Delta^{-1}D\Big\} = \mathrm{tr}\Big\{\Big(JXJ' - W_n^{-1}\Big)D\Big\} = \mathrm{tr}\Big\{XJ'DJ\Big\}, \tag{S22b}$$

$$\mathrm{tr}\Big\{11'\Delta^{-1}D\Delta^{-1}\Big\} = \Big(c'X'J' - v'J'\Big)D\Big(JXc - Jv\Big) = \Big(Xc - v\Big)'J'DJ\Big(Xc - v\Big). \tag{S22c}$$

The term in red, $J'DJ$, is a known matrix of order o, where o is the number of observed/sampled demes, and it can be precomputed and stored for easy access. Thus we do not need to construct the $n \times n$ matrix Δ^{-1} in order to compute $\mathrm{tr}\{\Delta^{-1}QD\}$; we can work with the $o \times o$ matrix X instead.

Next we show how to compute the pseudo determinant $\mathrm{Det}\{-\Delta^{-1}Q\}$. Following [Verbyla, 1990], we can show that

$$\mathrm{Det}\Big\{-\Delta^{-1}Q\Big\} = \frac{\det\{LL'\}}{\det\{-L\Delta L'\}} = \frac{(1'1)/(1'\Delta^{-1}1)}{-\det\{-\Delta\}}. \tag{S23}$$

A distance matrix is conditionally negative definite, and so Δ has one positive eigenvalue and $n-1$ negative eigenvalues [Bapat and Raghavan, 1997]. This guarantees that $-\det\{-\Delta\}$ is positive and so it is sufficient to compute $|\det\{\Delta\}| = |JBJ' - W_n|$, which can be obtained from the LU decomposition of $Y = BC - W_o$:

$$\Big|\det\{\Delta\}\Big| = \Big|\det\{W_n\}(-1)^{n-o}\det\{W_o^{-1}BC - I\}\Big| = \Big|\det\{W_n\}\det\{W_o^{-1}\}\det\{BC - W_o\}\Big|. \tag{S24}$$

We can use equations (S22), (S23) and (S24) to evaluate the log likelihood for the parameters k, m, q and σ^2, under the EEMS model (S13).

S1.4 Computing the resistance distances

Let $\mathcal{L}$ be the graph Laplacian of the population graph (V, E) with effective migration rates m. (The resistance distances do not depend on the diversity rates q, so those parameters are not relevant for the following computation.) The graph Laplacian is given by:

$$\mathcal{L} = \mathcal{D} - M, \tag{S25}$$

where $M = (m_{\alpha\beta})$ is the (sparse) matrix of effective migration rates and $\mathcal{D} = (\mathcal{D}_{\alpha\alpha})$ is the diagonal matrix with $\mathcal{D}_{\alpha\alpha} = \sum_{\beta:\beta\neq\alpha} m_{\alpha\beta}$.

Following [Babić et al., 2002], we can use $\mathcal{L}$ to compute the effective resistances R by inverting the sum matrix $\Gamma = \mathcal{L} + 11'/c$ where $c > 0$ is a constant. Let $H = \Gamma^{-1}$ and $h = \mathrm{diag}\{H\}$. Then

$$R = 1h' + h1' - 2H. \tag{S26}$$

Importantly, equation (S14) suggests that we can take $B = -2H$ instead of $B = R$ because, under the EEMS model (S13), the matrices $-2H$ and R give the same likelihood and are therefore equivalent. In practice, this means that we can avoid inverting the $d \times d$ matrix Γ to obtain the auxiliary matrix X in (S18). Let Γ_{11} be the $o \times o$ block that corresponds to the observed demes; similarly, let Γ_{22} be the $(n-o) \times (n-o)$ block that corresponds to the unobserved demes. Then

$$H_{11}^{-1} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}, \tag{S27}$$

which can be computed efficiently by solving a linear system. Finally, the dissimilarities B_{11} between observed demes can be computed from $B_{11}^{-1} = -H_{11}^{-1}/2$. If the population graph G is sparsely sampled, as is often the case, it is more efficient to compute the Schur complement of Γ_{11} given by equation (S27), rather than invert the full-size matrix Γ. This idea is also used in [Hanks and Hooten, 2013].

S1.4.1 Computational complexity

The auxiliary matrix Γ is dense, diagonally dominant, positive definite and of order d, where d is the number of observed demes. First we compute the Schur complement H_{11}^{-11} according to equation (S27), and then the between-demes dissimilarities B_{11}.

1. Cholesky decomposition $\Gamma_{22} = U'U$: $O((d-o)^3)$
2. Forward substitution $U'Y = \Gamma_{21}$: $O(o(d-o)^2)$
3. Backward substitution $UX = Y$: $O(o(d-o)^2)$
4. Matrix inversion $B_{11} = -H_{11}^{-1}/2$: $O(o^3)$

This procedure has complexity $O((d+o)(d-o)^2 + o^3)$ and, expect for very small graphs G, it is more efficient than inverting the sum matrix Γ which has complexity $O(d^3)$.

S1.5 Markov Chain Monte Carlo estimation

EEMS makes use of two Voronoi tessellations, which independently partition the habitat: one parametrizes the effective migration rates m, and the other – the effective diversity rates q. Specifically, the migration rates m are determined by a Voronoi tessellation with C_m cells, seeds $s_1, \ldots, s_{C_m}$, migration effects $e_1, \ldots, e_{C_m}$, and overall migration rate (on the $\log_{10}$ scale) μ, while the diversity rates are determined by another independent Voronoi tessellation with C_q cells, seeds $t_1, \ldots, t_{C_q}$, and diversity effects $f_1, \ldots, f_{C_q}$. The overall diversity rate (on the $\log_{10}$ scale) is assumed to be 0 because, following eq. (S15), $\mathrm{E}\{D \mid m, q, \sigma^2\} = \sigma^2\Delta(m,q) = \sigma^2\Big(JB(m)J' - W(q)\Big)$, where $B(m)$ and $W(q)$ are the between-demes and the within-deme components of the expected dissimilarity, and they scale so that $B(m/2) = 2B(m)$, $W(2q) = 2W(q)$. With this parametrization, fixing the overall diversity rate to 1 makes the scale σ^2 identifiable. Finally, the migration cell effects $e_1, \ldots, e_{C_m}$, have variance ω_m^2, while the diversity cell effects $f_1, \ldots, f_{C_q}$ have variance ω_q^2.

We use birth-death MCMC to estimate the number of cells C in each Voronoi tessellation because the dimension of the seeds and the cell effects changes as C increases or decreases. In each step, we propose the birth (addition) or death (removal) of a cell, with equal probability [Stephens, 2000]. For a birth proposal, the acceptance probability is

$$\alpha(\Theta, \Theta^*) = \min\left\{1, u\frac{c+r}{c+1}\frac{\ell(\Theta^*)}{\ell(\Theta)}\right\}, \tag{S28}$$

where c is the current number of cells; (r,u) are the parameters of the negative binomial prior on C, Θ is the current parameter state (with c cells) and Θ^* is the proposed parameter state, with one additional cell added at a random location s^* within the habitat and assigned migration effect e^* drawn from a (truncated) normal

prior. Small probability of success u means small acceptance probability α unless the likelihood ratio indicates strong evidence in favor of adding the new cell.

For a death proposal, one cell is randomly chosen to be removed and the acceptance probability has a very similar form:

$$\alpha(\Theta, \Theta^*) = \min\left\{1, \frac{1}{u}\frac{c+1}{c+r}\frac{\ell(\Theta^*)}{\ell(\Theta)}\right\}. \tag{S29}$$

For SNP data, the parameter state $\Theta = (k, m, q, \sigma^2)$ consists of the degrees of freedom k, the migration rates m, the diversity rates q and a scale parameter σ^2; the likelihood $\ell(\Theta)$ is given by equation (3). For microsatellite data, the parameter state $\Theta = (m, q, \sigma_1^2, \ldots, \sigma_p^2)$ consists of the migration rates m, the diversity rates q and locus-specific scale parameters $\sigma_1^2, \ldots, \sigma_p^2$; the likelihood $\ell(\Theta)$ is given by equation (7).

For a given number of Voronoi cells, C_m and C_q, in the two Voronoi tessellations, the cell effects and their locations, the overall migration rate μ and (for SNPs data only) the effective degrees of freedom k are each updated in turn with a random-walk Metropolis-Hastings step:

$$\alpha(\Theta, \Theta^*) = \min\left\{1, \frac{p(\Theta^*)}{p(\Theta)}\frac{\ell(\Theta^*)}{\ell(\Theta)}\right\}, \tag{S30}$$

where $p(\Theta)$ is the prior and $\ell(\Theta)$ is the likelihood.

Finally, the scalar variance parameters are $\omega_m^2, \omega_q^2, \sigma^2$ for SNP data, and $\omega_m^2, \omega_q^2\sigma_1^2, \ldots, \sigma_p^2$ for microsatellite data. These parameters are updated with a Gibbs step by sampling from the corresponding full conditional distribution, which is inverse gamma. For example, the variance in relative migration among cells, ω_m^2, is drawn from

$$\omega_m^2 \,|\, C, e_1 \ldots, e_C \sim \text{Inv-G}\Big((c_\omega + C)/2, (d_\omega + SS_e)/2\Big), \tag{S31}$$

where the sum of squares is $SS_e = \sum_{c=1}^{C} e_c$.

S1.6 Color scheme

To simplify comparisons, we plot relative migration rates on the same scale throughout the paper; the blue-and-orange color palette is based a collection of divergent color schemes suitable for people with deficient red-green vision. The choice of color and scaling can, as with any plot of this nature, affect – sometimes profoundly – the resulting image and the message it conveys. At the simplest level, a scheme that is too broad will wash-out any differences in effective migration rate between regions, while a scheme that is too narrow will risk over-emphasizing trivial differences. The color scheme used here, as well as some alternatives, is available at `http://geog.uoregon.edu/datagraphics/color_scales.htm` [Light and Bartlein, 2004].

An estimated effective migration surface (EEMS) is a colored contour plot. To create the plot:

1. Choose a dense grid of interpolation points (x, y) for the contour plot. The interpolation points do not correspond to the demes in the population graph, and we would choose a very dense grid of interpolation points to obtain a smooth migration surface.

2. Compute the effective migration rate, $\log_{10}(m_{xy})$, at each interpolation point, from the estimated Voronoi tessellation of the effective migration rates. The Voronoi tessellation partitions the habitat, and each interpolation point is assigned the rate of the cell c which it falls into, $\log_{10}(m_c)$.

3. Standardize the migration rates, so that the mean over the interpolation points is 0.

The Voronoi cells do not necessarily have the same area, and so different number of interpolation points would fall in each cell. Therefore, the unweighted average over the interpolation points corresponds to a weighted average across the Voronoi cells, with weights proportional to the area of each cell. This is the reason we do not use the estimate of the overall migration μ to standardize the interpolated rates for plotting. (Recall that $\log_{10}(m_c) = \mu + e_c$ for cell c, and this implies μ is the unweighted mean over the cells. However, for plotting an estimated effective migration surface, we want the "average color" across the habitat to be white.)

References

[Babić et al., 2002] Babić, D., Klein, D. J., Lukovits, I., Nikolić, S., and Trinajstić, N. (2002). Resistance-distance matrix: a computational algorithm and its application. *Int. J. Quantum. Chem.*, 90:166–176.

[Bapat and Raghavan, 1997] Bapat, R. B. and Raghavan, T. E. S. (1997). *Nonnegative matrices and applications*. Cambridge University Press, Cambridge, UK.

[Guillot et al., 2005] Guillot, G., Estoup, A., Mortier, F., and Cosson, J. F. (2005). A spatial statistical model for landscape genetics. *Genetics*, 170:1261–1280.

[Hanks and Hooten, 2013] Hanks, E. M. and Hooten, M. B. (2013). Circuit theory and model-based inference for landscape connectivity. *J. Am. Stat. Assoc.*, 108:22–33.

[Henn et al., 2011] Henn, B. M., Gignoux, C. R., Jobin, M., Granka, J. M., Macpherson, J. M., Kidd, J. M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., Lin, A. A., Underhill, P. A., Comas, D., Kidd, K. K., Norman, P. J., Parham, P., Bustamante, C. D., Mountain, J. L., and Feldman, M. W. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl. Acad. Sci. U.S.A.*, 108:5154–5162.

[Hubisz et al., 2009] Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.*, 9:1322–1332.

[Hudson, 1990] Hudson, R. R. (1990). Gene genealogies and the coalescent process. In Futuyma, D. and Antonovics, J., editors, *Oxford surveys in evolutionary biology*, volume 7, pages 1–44. Oxford University Press.

[Kimura and Weiss, 1964] Kimura, M. and Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49:561–576.

[Light and Bartlein, 2004] Light, A. and Bartlein, P. J. (2004). The end of the rainbow? Color schemes for improved data graphics. *Eos*, 85:385.

[McVean, 2009] McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genet.*, 5:e1000686.

[Ohta and Kimura, 1973] Ohta, T. and Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.*, 22:201–204.

[Platt et al., 2010] Platt, A., Horton, M., Huang, Y. S., Li, Y., Anastasio, A. E., Mulyati, N. W., Ågren, J., Bossdorf, O., Byers, D., Donohue, K., Dunning, M., Holub, E. B., Hudson, A., Corre, V. L., Loudet, O., Roux, F., Warthmann, N., Weigel, D., Rivero, L., Scholl, R., Nordborg, M., Bergelson, J., and Borevitz, J. O. (2010). The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.*, 6:e1000843.

[Pritchard et al., 2000] Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959.

[Stephens, 2000] Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components —an alternative to reversible jump methods. *Ann. Stat.*, 28:40–74.

[Verbyla, 1990] Verbyla, A. P. (1990). A conditional derivation of residual maximum likelihood. *Aust. J. Stat.*, 32:227–230.

[Wasser et al., 2014] Wasser, S. K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C., and Weir, B. (2014). Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. *Submitted*.

[Wasser et al., 2004] Wasser, S. K., Shedlock, A. M., Comstock, K., Ostrander, E. A., Mutayoba, B., and Stephens, M. (2004). Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade. *Proc. Natl. Acad. Sci. U.S.A.*, 10:14847–14852.

S2 Supplementary Figures

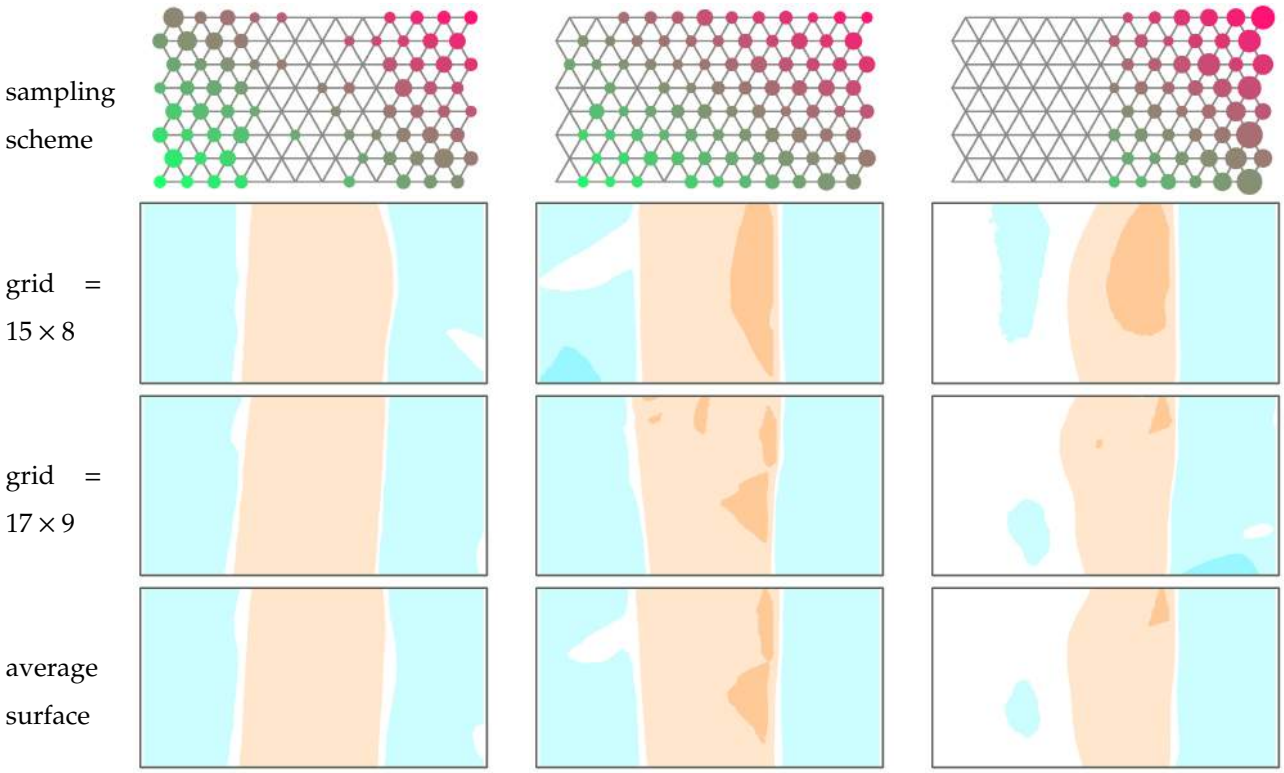


Figure S1 The effect of the grid choice on the inferred EEMS, using the barrier-to-migration simulation in Figure 2 for illustration. The population grid $G = (V, E)$ is fixed (the vertices V and the edges E are specified as input arguments) and it can affect inference. For example, a very coarse grid might not allow for sufficient variation in resistance distances. Therefore, we recommend averaging the EEMS inferred for grids of different sizes – in practice this will moderate the discretization implicit in assigning the samples collected in a continuous habitat to the vertices of a discrete population graph. The data for this example was simulated on a 12×8 grid but we use 15×8 and 17×9 grids to estimate the EEMS. As the grid dimensions change, so do the occupied demes because samples are assigned to the closest deme in the grid. When averaging surfaces with grids of different size, the broad features remain the same but small details can change.

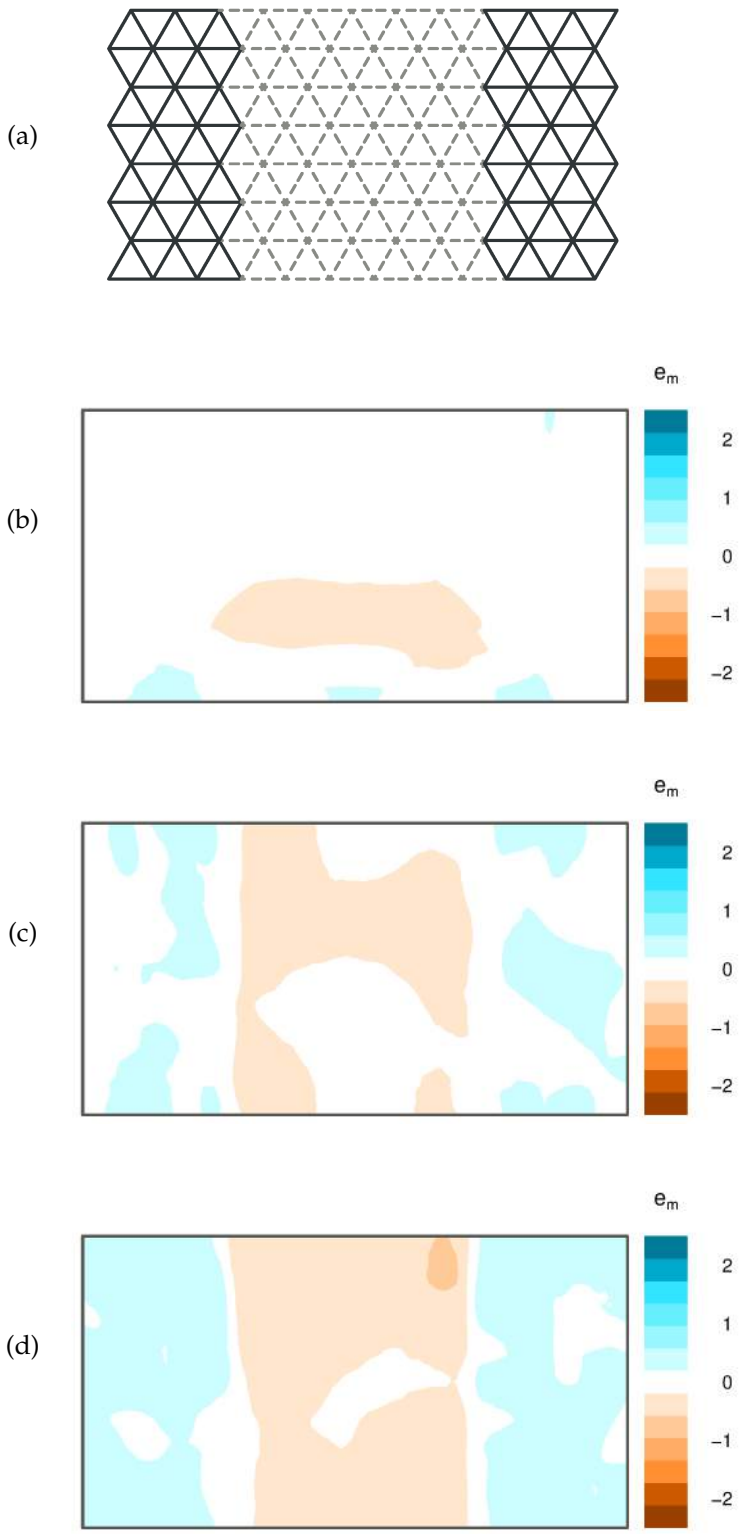


Figure S2 A past demographic event produces a barrier to effective migration. **(a)** An ancestral population splits into subpopulations E (east) and W (west), separated by an "effective barrier" to migration: an area where the migration rates drop simultaneously to 0 at a point of time, x, in the past. The further back in time the split event occurs, the more differentiated subpopulations E and W are. The split occurs at **(b)** $x = 1$; **(c)** $x = 4$; **(d)** $x = 9$ units of time in the past, which is measured in N_0 generations. Before the split, migration rates are all set to 1 (on the same coalescent scale, N_0); after the split, migration rates on either side of the central region remain 1.

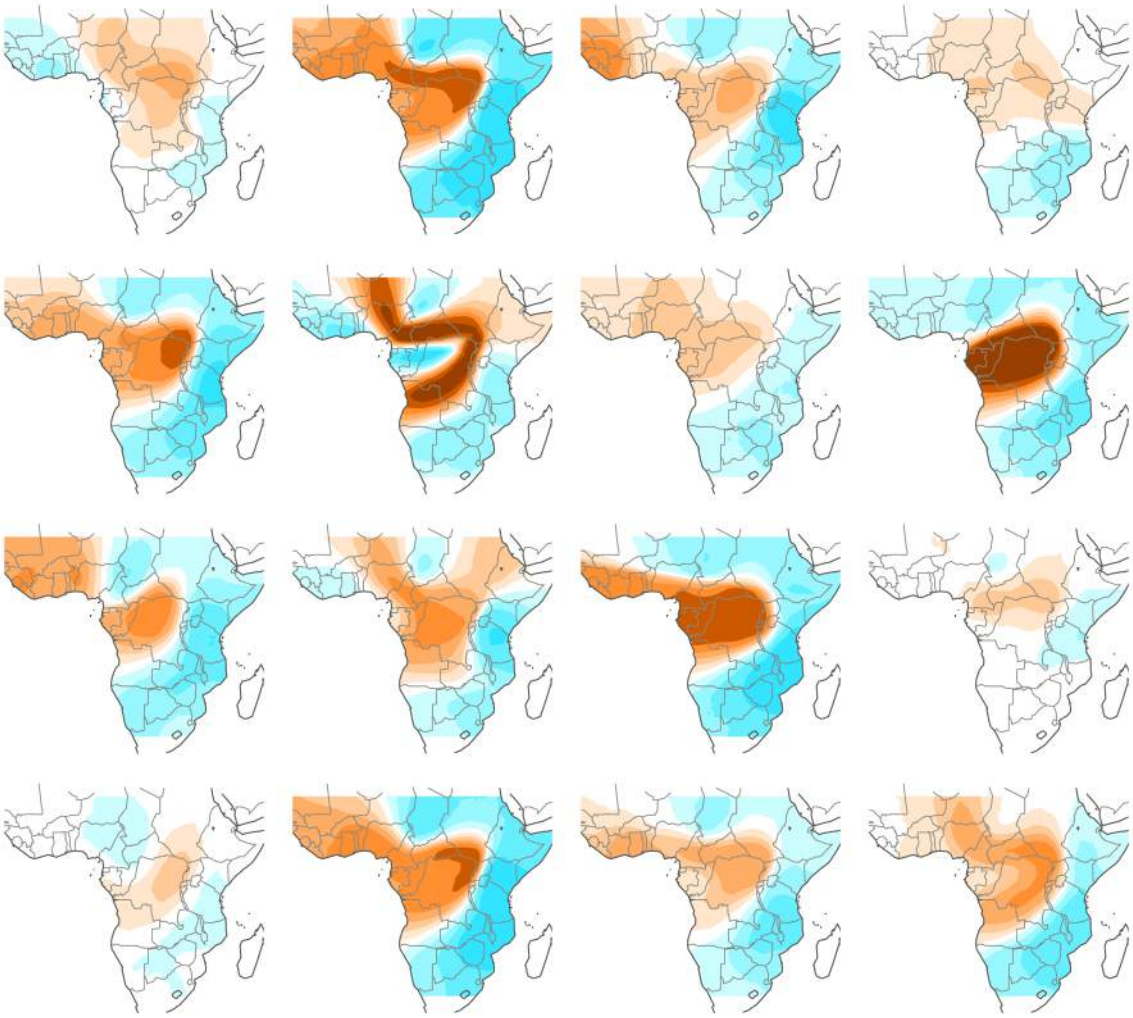
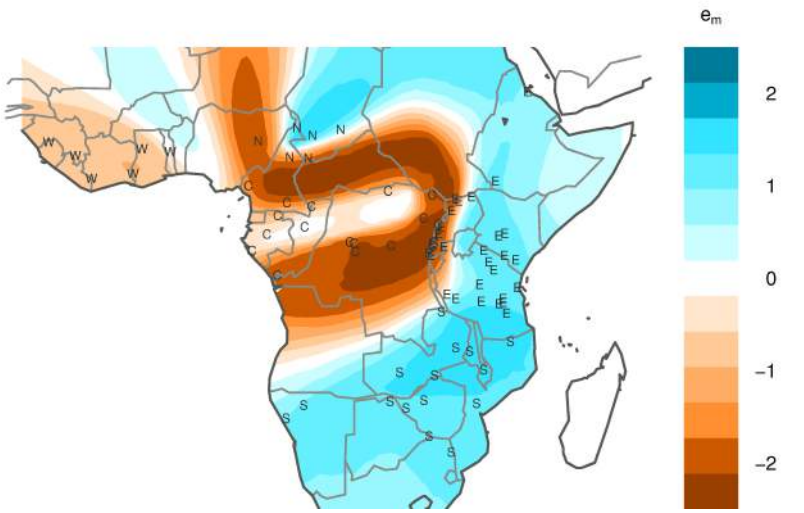
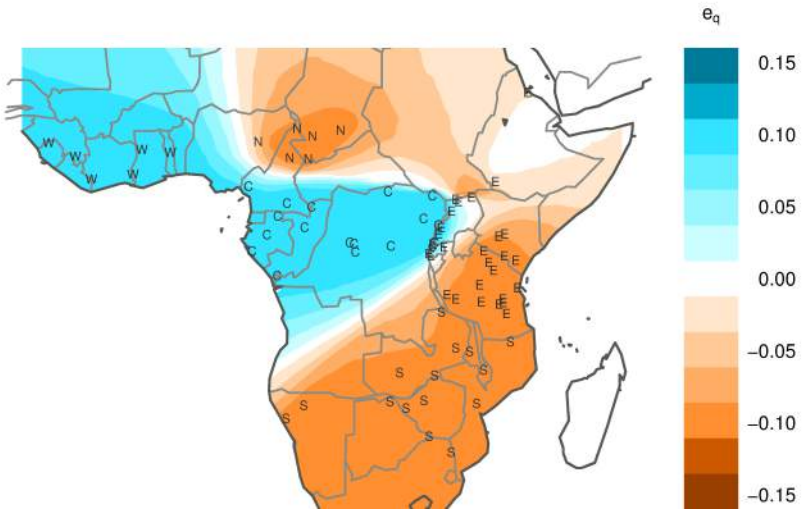


Figure S3 Effective migration rates for the African elephant at sixteen microsatellite loci. The loci are highly polymorphic and therefore informative for the sample genealogy at each site: every mutation on the genealogy contains information about the branch lengths in the tree (since longer branches are more likely to carry a mutation, if the mutation rate is constant in time). The sixth locus is extremely informative, presumably because it has the highest mutation rate, and thus it successfully captures the strong effective barrier to migration between the habitat ranges of forest and savanna elephants.



(a)



(b)

Figure S4 Further EEMS analysis of the population structure of the African elephant (data from [Wasser et al., 2014]). **(a)** Estimated effective migration rates, after excluding the most variable locus (the sixth locus in Supplementary Figure S3). **(b)** Estimated effective diversity rates.

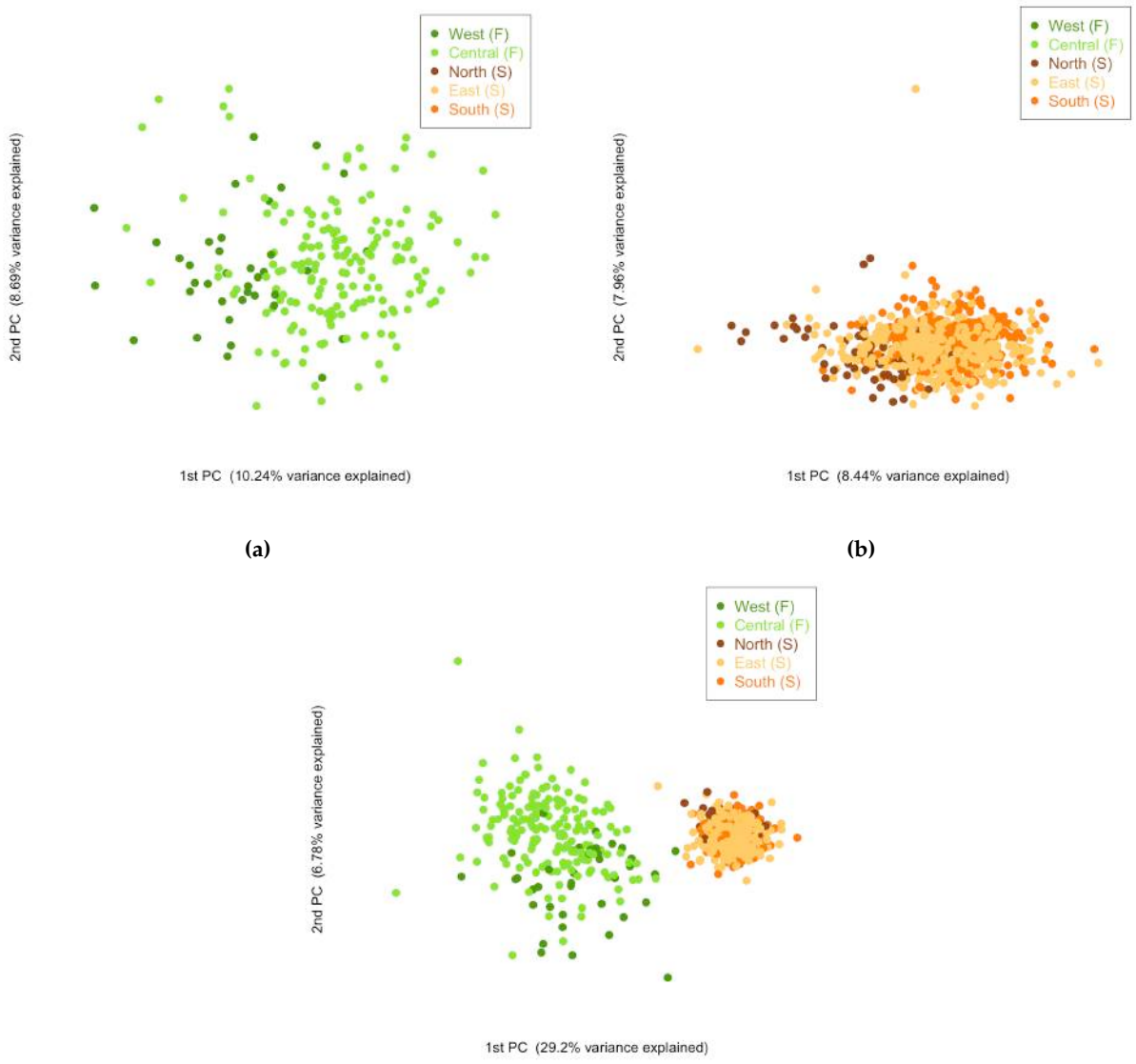


Figure S5 Principal component analysis of the African elephant data: **(a)** forest elephants, from the West and Central regions; **(b)** savanna elephants, from the North, East and South regions; **(c)** forest and savanna elephants.

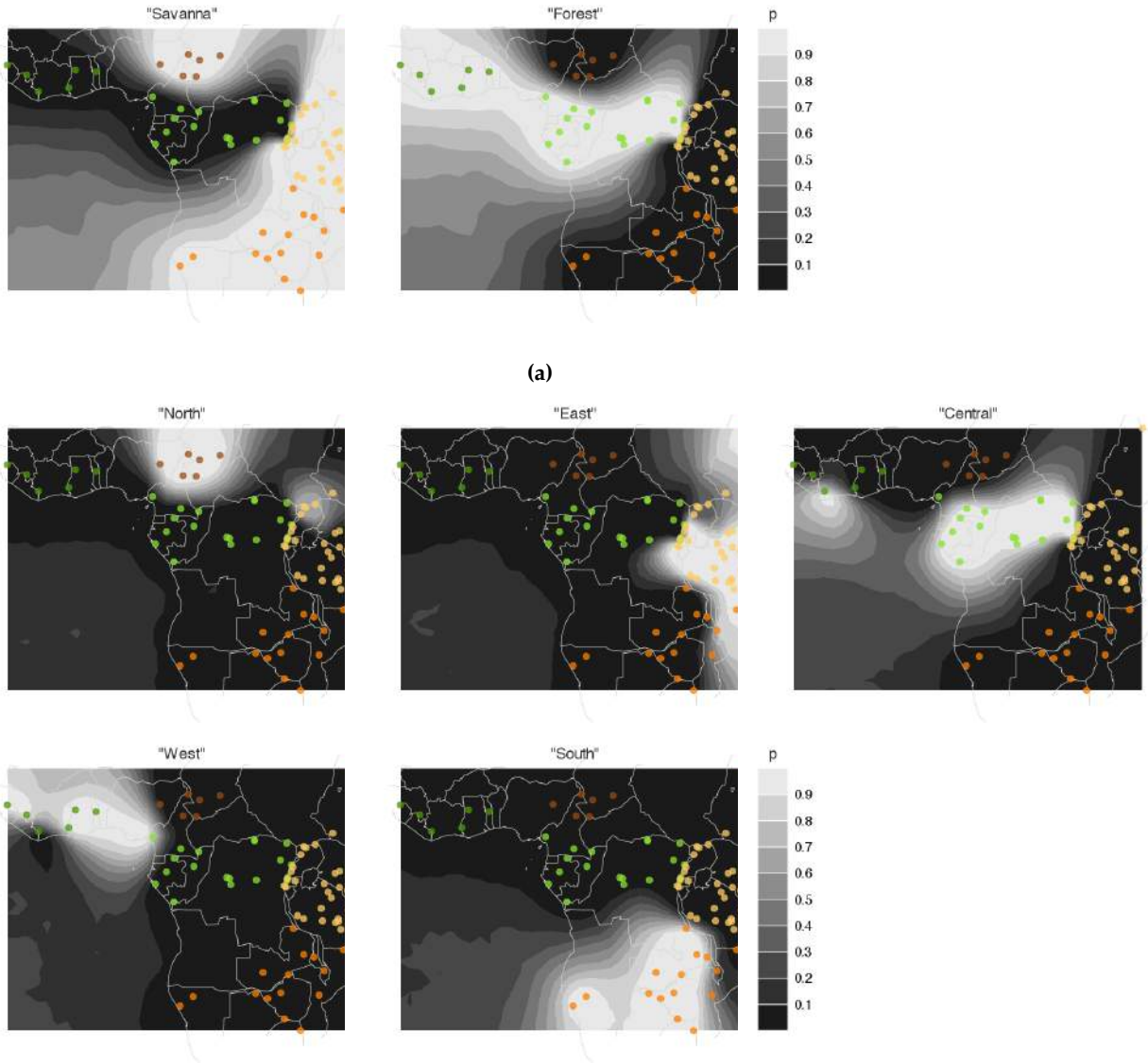
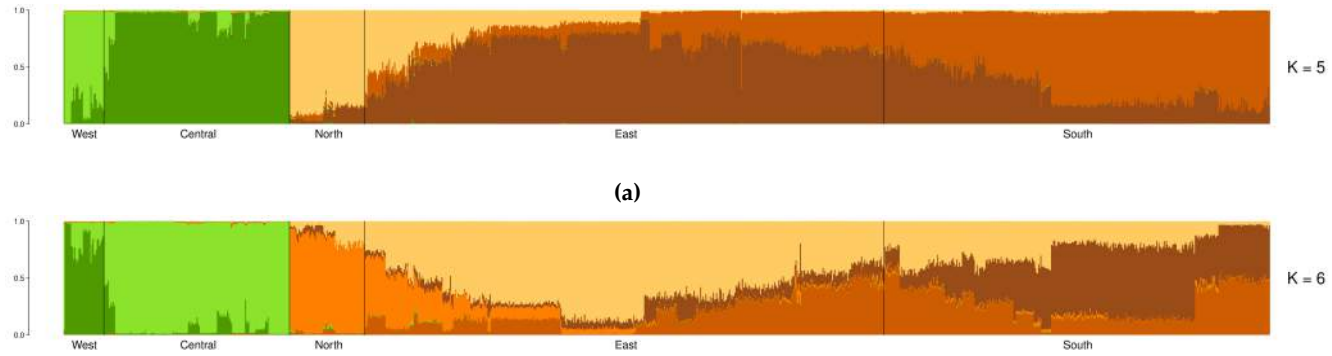


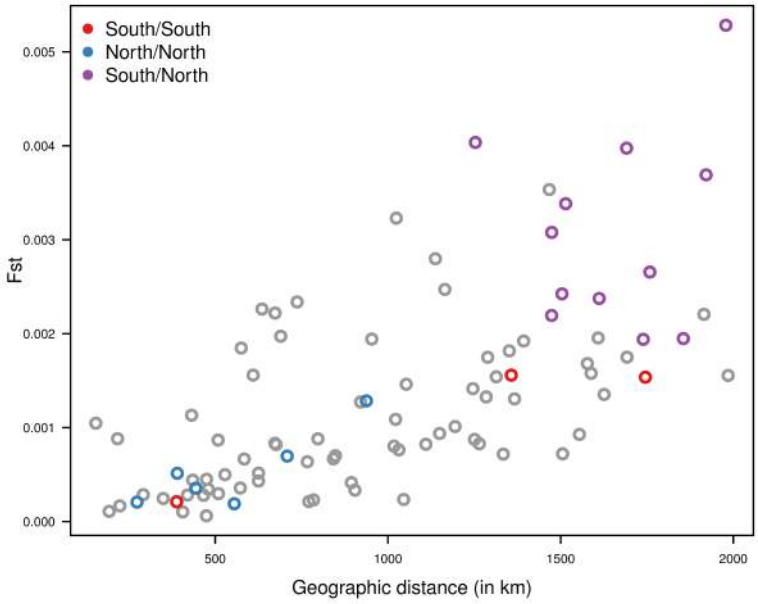
Figure S6 GENELAND analysis of the African elephant data. GENELAND [Guillot et al., 2005] is a cluster-based method which uses a Voronoi tessellation to encourage spatially continuous clusters and find sharp boundaries between genetically differentiated groups. **(a)** Posterior probabilities for belonging to each of two inferred clusters, which correspond to the ranges of the savanna and forest subspecies. **(b)** Posterior probabilities for belonging to each of five inferred clusters, which correspond to the five biogeographic regions defined in [Wasser et al., 2004]: "North", "South", "East", "West" and "Central". GENELAND successfully detects differences in allele frequencies between the two species and the five biogeographic regions. However, GENELAND does not model the relationships between the regions: the five clusters in **(b)** are as distinct from each other as the two clusters in (a), even though the "West" and "Central" clusters are inhabited by forest elephants while the "North", "East" and "South" clusters – by savanna elephants.



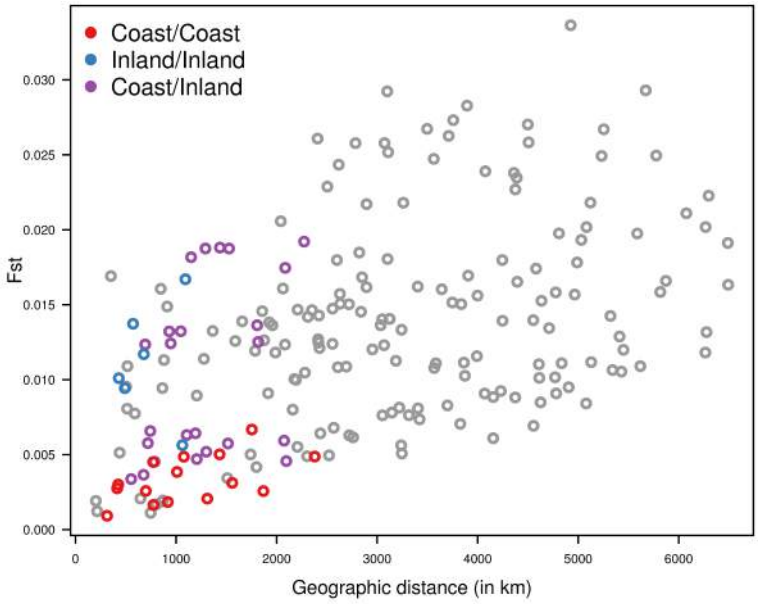
(a)

(b)

Figure S7 STRUCTURE analysis of the African elephant data. **(a)** Membership proportions for belonging to five inferred clusters. **(b)** Membership proportions for belonging to six inferred clusters. [Individuals are ordered by sampling location.] Compared to GENELAND [Guillot et al., 2005], STRUCTURE [Pritchard et al., 2000] with a sampling location prior [Hubisz et al., 2009] provides intuition for the relationship between the five biogeographic regions. STRUCTURE clearly detects the difference between forest elephants (West and Central regions) and savanna elephants (North, East and South regions) as they fall into different clusters. Furthermore, STRUCTURE shows some evidence for isolation by distance, particularly in savanna elephants, as most of these individuals are represented as weighted mixtures of ancestral clusters that do not correspond to distinct geographic areas.

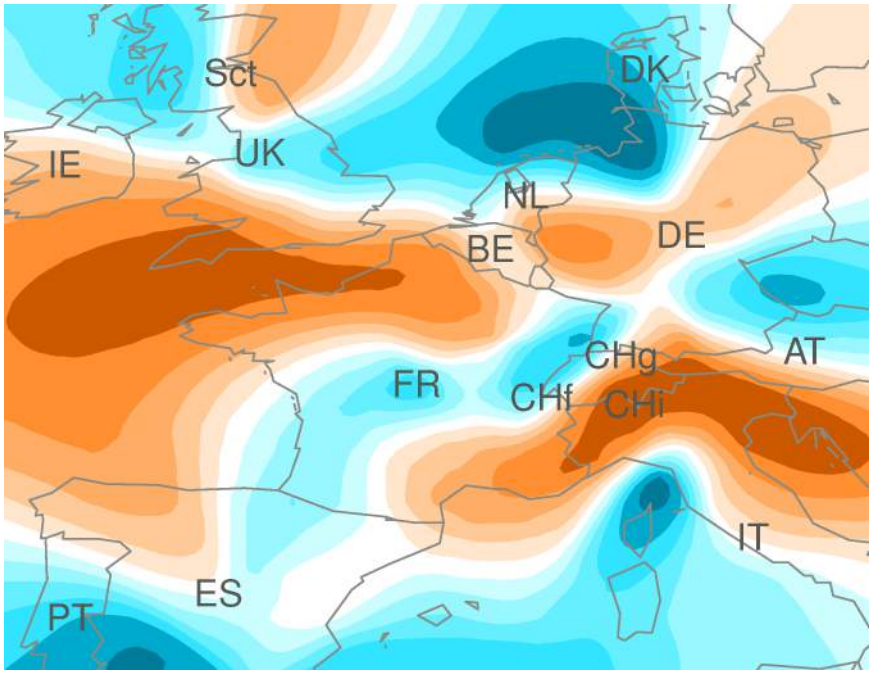


(a)

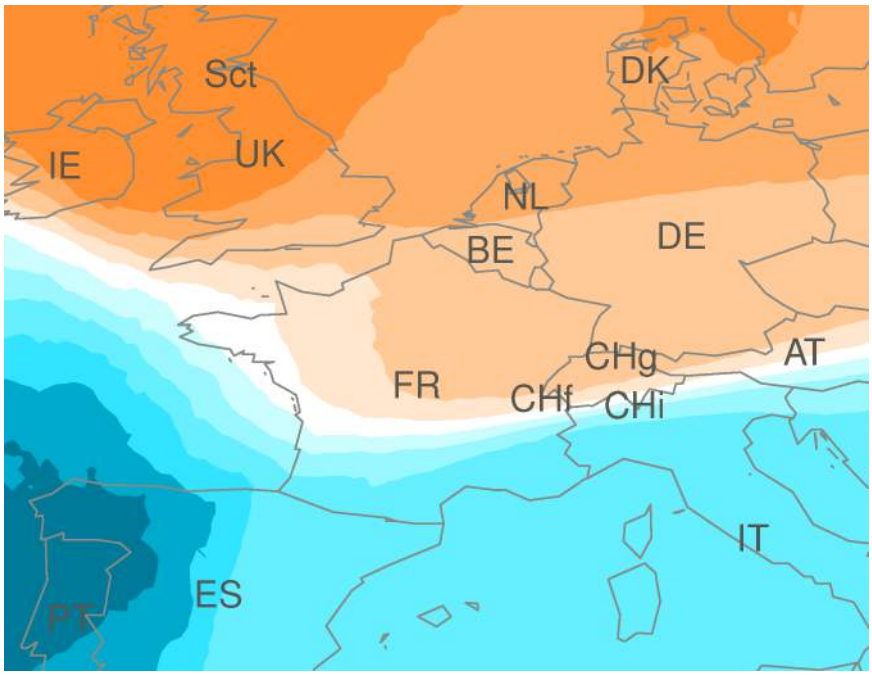


(b)

Figure S8 Genetic dissimilarity (F_{ST}) as a function of geographic (great circle) distance, for European and African populations. For both datasets, differentiation increases with distance, which suggests that spatial structure is consistent with isolation by distance. The colors are chosen to emphasize comparisons between two groups of populations. **(a)** In Western Europe, the "south" group consists of Portugal (PT), Spain (ES), Italy (IT); the "north" group consists of Ireland (IE), Scotland (Sct), United Kingdom (UK), Holland (NL). Comparisons within the "south" and "north" groups are colored red and blue, respectively; comparisons between the two groups are colored purple. There is greater similarity within each group than between the groups. **(b)** In Sub-Saharan Africa, the "coast" group consists of Brong (Br), Yoruba (Yo), Igbo (Ig), Bamoun (Ba2), Fang (Fa), Kongo (Ko); the "inland" group consists of Hausa (Ha), Mada (Ma2), Bulala (Bu), Kaba (Ka). Coastal populations are more similar genetically than inland populations, even though some coastal populations are further apart.

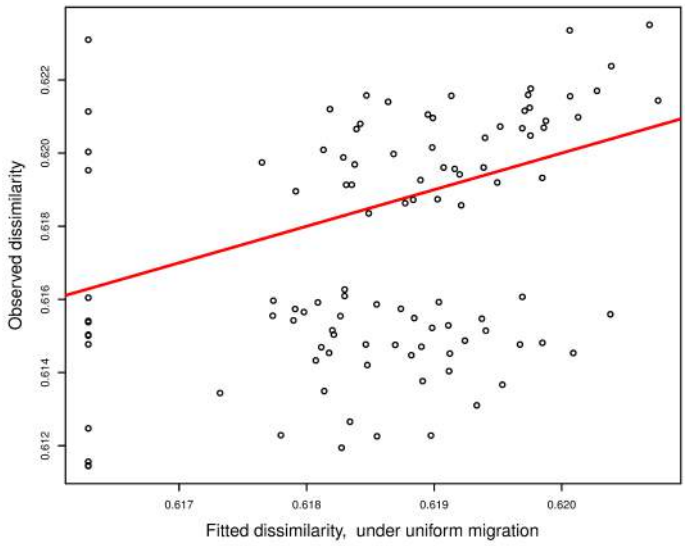


(a)

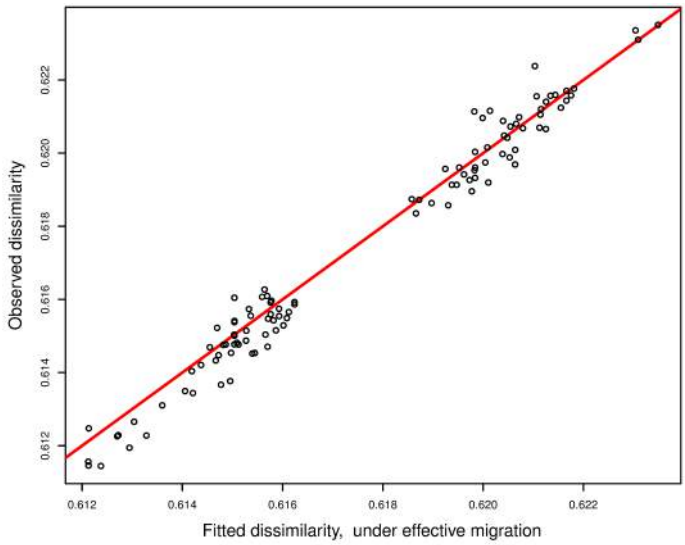


(b)

Figure S9 EEMS analysis of the spatial structure in Western European populations. EEMS estimates the effective diversity rates within demes, jointly with the effective migration rates between connected demes. The fitted diversity rates can be interpolated across the habitat to produce an "estimated effective diversity surface", which is complementary to the "estimated effective migration surface". **(a)** Effective migration rates. **(b)** Effective diversity rates. This plot highlights the previously noted north-south gradient in human diversity in Europe.

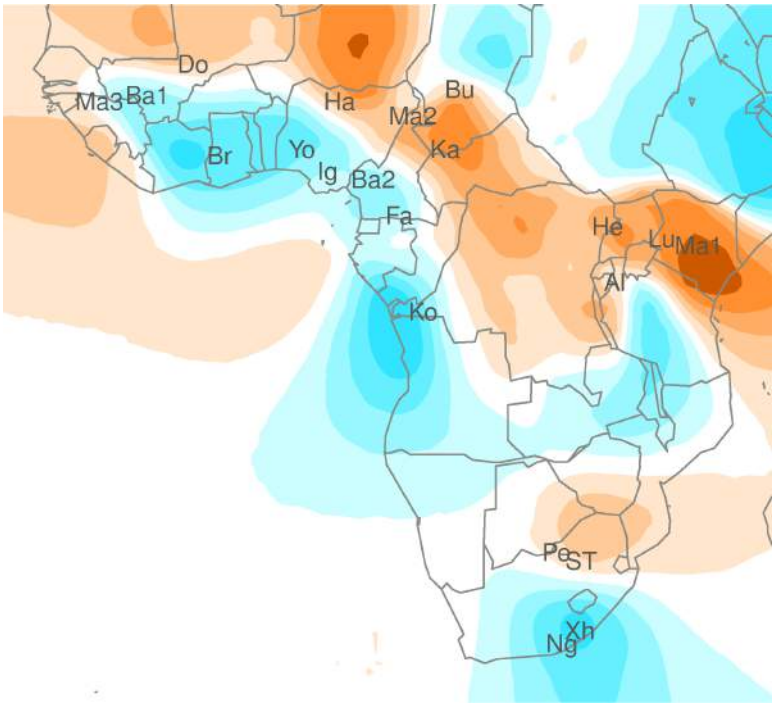


(a)

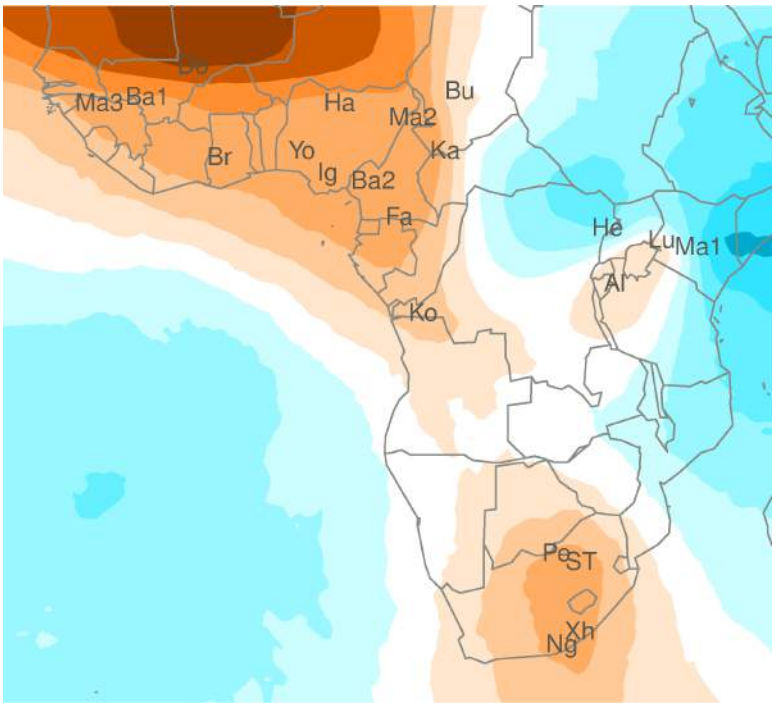


(b)

Figure S10 Observed versus fitted dissimilarities for 14 Western European populations: Austria (AT), Belgium (BE), Switzerland (samples are categorized further as French, German and Italian speaking Swiss: CHf, CHg, CHi), Germany (DE), Spain (ES), France (FR), Ireland (IE), Italy (IT), Netherlands (NL), Portugal (PT), Scotland (Sct), United Kingdom (UK). We have excluded Denmark (DK) from this scatterplot (although not from EEMS) because there is a single sampled individual from this population, and so the observed genetic dissimilarity is 0. Genetic dissimilarities between pairs of individuals are averaged according to their assignment on a regular 19×15 population grid. **(a)** Dissimilarities are inferred under uniform migration, a setting that simulates exact isolation by distance. **(b)** Dissimilarities are inferred with EEMS, which estimates both the effective migration rates and the effective diversity rates, assuming equilibrium in time. Since the EEMS model specifies that the fitted genetic distances predict the observed genetic differences, we can use the coefficient of determination, r^2, to assess the goodness-of-fit: **(a)** $r^2 = 0.142$; **(b)** $r^2 = 0.978$.

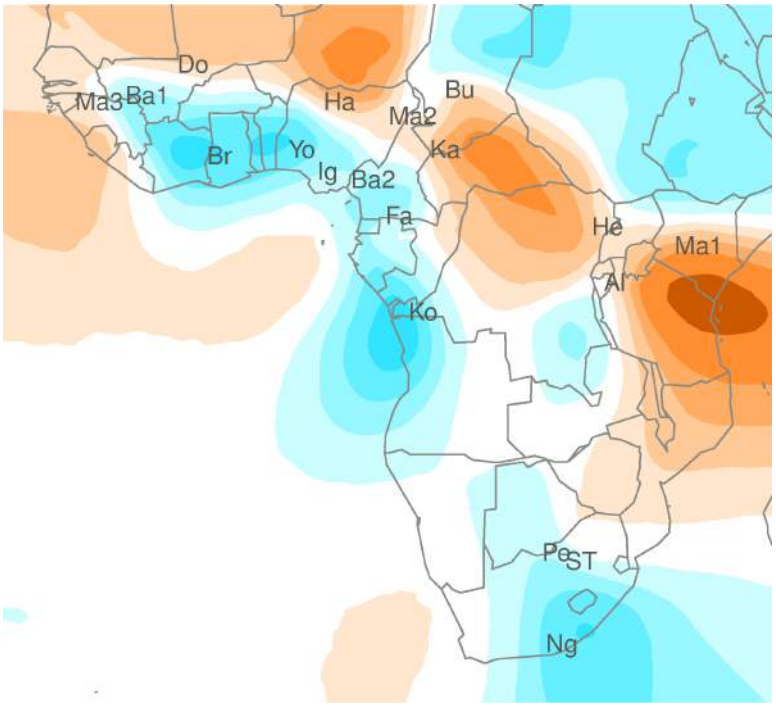


(a)

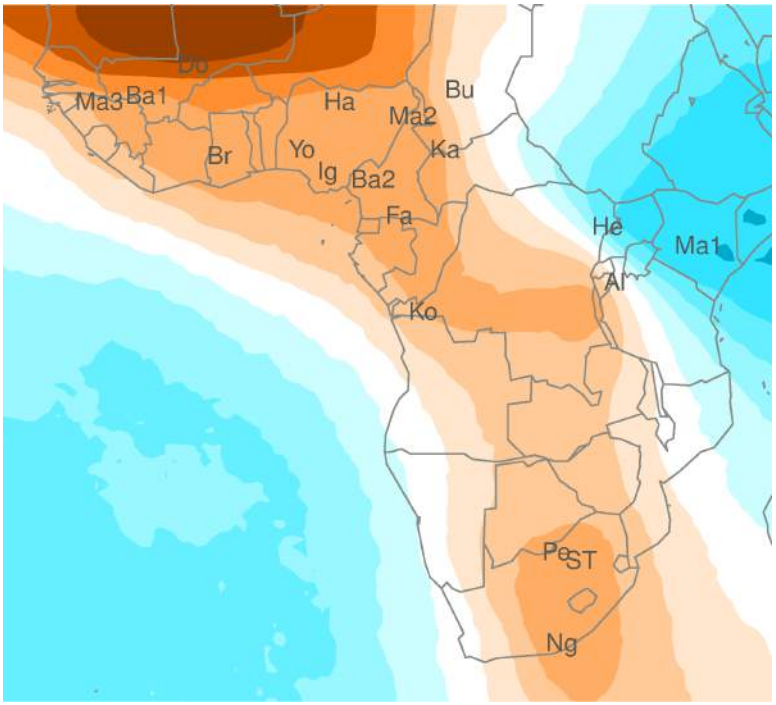


(b)

Figure S11 EEMS analysis of the spatial structure in 21 Sub-Saharan populations: Alur (Al), Bambaran (Ba1), Bamoun (Ba2), Brong (Br), Bulala (Bu), Dogon (Do), Fang (Fa), Hausa (11), Hema (He), Igbo (Ig), Kaba (17), Kongo (Ko), Luhya (Lu), Maasai (Ma1), Mada (Ma2), Mandenka (Ma3), Nguni (Ng), Pedi (Pe), Sotho/Tswana (ST), Xhosa (Xh), Yoruba (Yo). **(a)** Effective migration rates. **(b)** Effective diversity rates. This plot highlights higher diversity in East Africa than in South or West Africa.

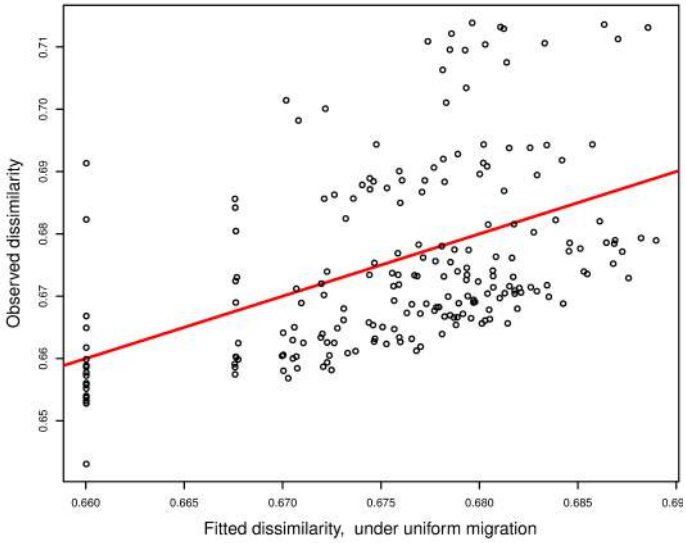


(a)

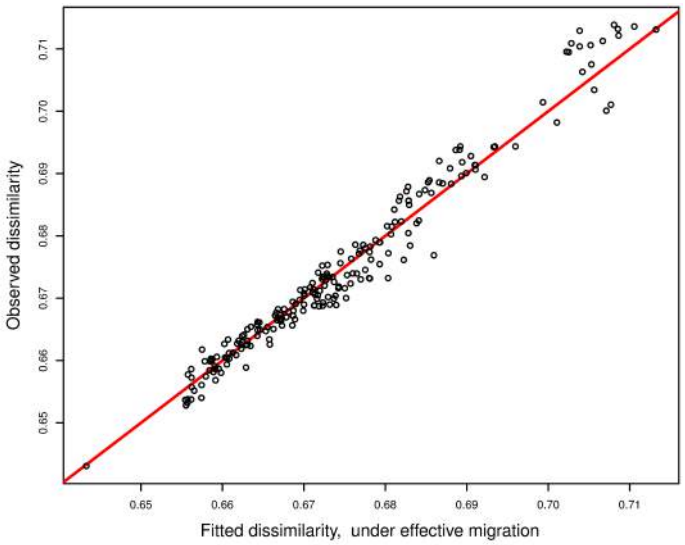


(b)

Figure S12 EEMS analysis of the spatial structure in 19 Sub-Saharan populations, after excluding the Luhya (Lu) and the Xhosa (Xh), two Bantu speaking populations considered recent geographic migrants [Henn et al., 2011]. **(a)** Effective migration rates. **(b)** Effective diversity rates.

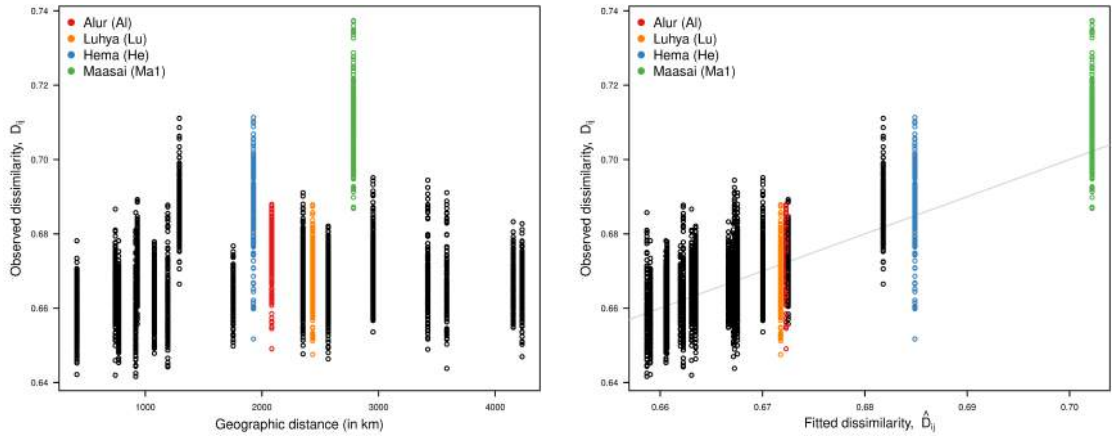


(a)

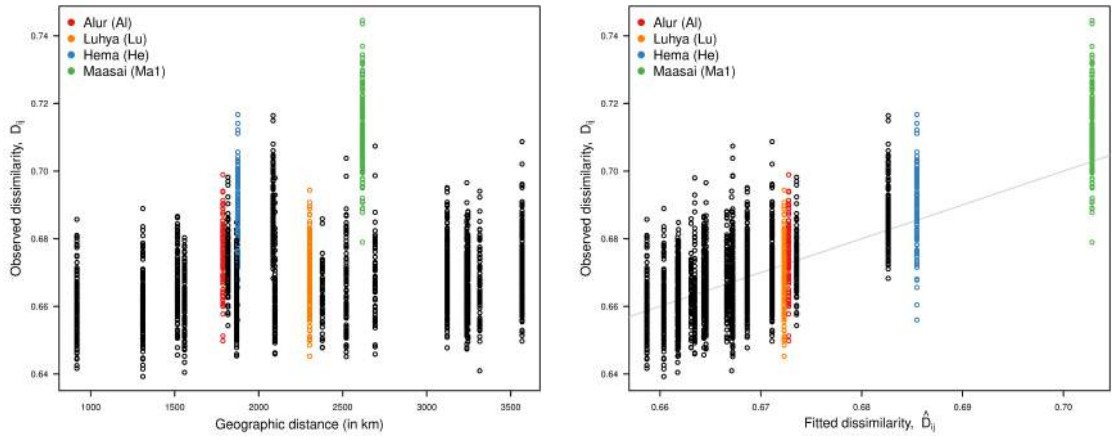


(b)

Figure S13 Observed versus fitted dissimilarities for Sub-Saharan African populations. Genetic dissimilarities between pairs of individuals are averaged according to their assignment on a regular 19×17 population grid. **(a)** Dissimilarities are inferred under uniform migration, which simulated exact isolation by distance; $r^2 = 0.164$. **(b)** Dissimilarities are inferred with EEMS; $r^2 = 0.914$, where r^2 is the proportion of variance in the observed genetic differences explained by the fitted genetic differences.

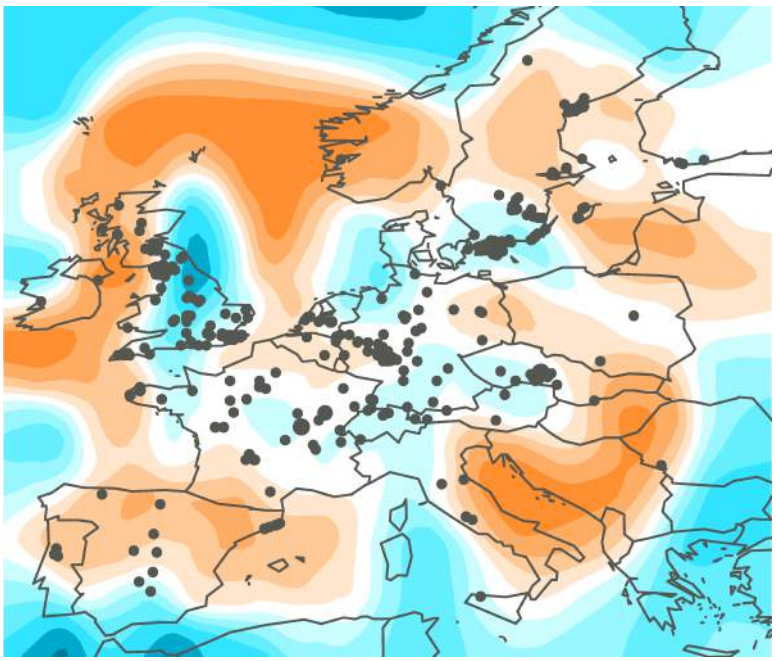


(a)

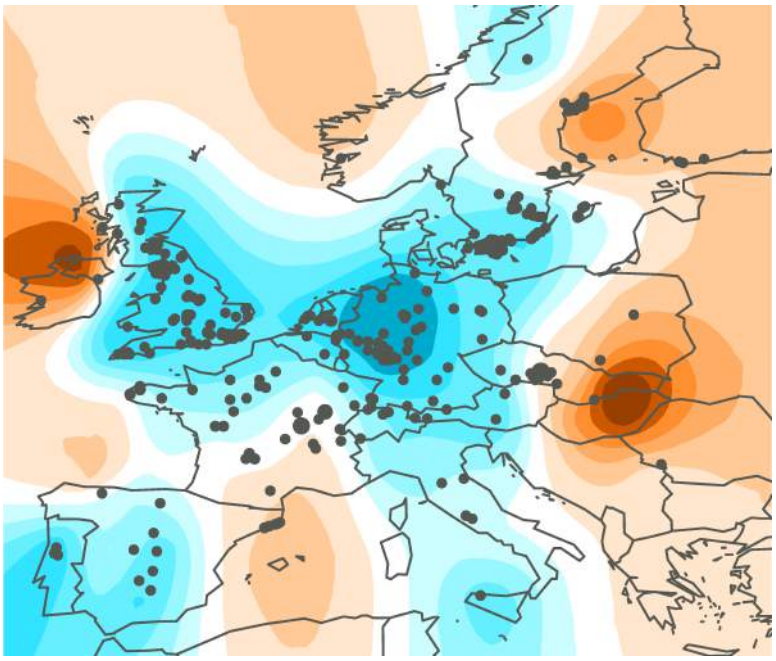


(b)

Figure S14 Observed genetic dissimilarities vs geographic distances and fitted dissimilarities between two Bantu speaking groups – the Fang (Fa) and the Kongo (Ko) – and other ethnic groups in Sub-Saharan Africa. **(a)** Dissimilarities between a Fang individual and an individual from another ethnic group. **(b)** Dissimilarities between a Kongo individual and an individual from another ethnic group. The Luhya (Lu) – a Bantu speaking group in the east – are further away from the Fang and the Kongo than the Alur (Al) but just as similar genetically.



(a)



(b)

Figure S15 EEMS analysis of 980 *Arabidopsis thaliana* accessions from Europe. **(a)** Effective migration surface. There is relatively little variation in migration rates across continental Europe (France, Germany, Central Europe), as expected under isolation by distance [Platt et al., 2010]. **(b)** Effective diversity surface. Diversity is highest in Germany and Central Europe, and tends to decrease in coastal regions and at the boundaries of the sampled habitat.