File S3

The MCMC sampling scheme

We used a random-walk Metropolis-Hastings algorithm to sample the joint posterior density of $\boldsymbol{q}$, $\boldsymbol{a}$ and $\boldsymbol{\kappa}$. Below, we describe how each parameter was sampled while keeping the other parameters fixed.

- Sampling the drift parameters $\boldsymbol{a}$. We used N($a_k, \delta_{a_k}^2$) distributions separately for each k to draw proposals for $\log a_k$. The variance parameters $\delta_{a_k}^2$ were adjusted during the burn-in as in Ovaskainen *et al*. (2008) to give an accept ratio of 0.44.
- Sampling lineage loadings $\boldsymbol{\kappa}$. We used TDD($\delta_{\boldsymbol{\kappa}_A}\boldsymbol{\kappa}_A$), i.e. truncated Dirichlet, distributions (Fang *et al*. 2000) separately for each A and j to draw proposals for $\boldsymbol{\kappa}_A$. The $\delta_{\boldsymbol{\kappa}_A}$'s are proposal parameters that were adjusted during the burn-in as in Ovaskainen *et al*. (2008) to give an accept ratio of 0.44.
- Sampling ancestral allele frequencies $\boldsymbol{q}$ and lineage-specific allele frequencies $\boldsymbol{z}$. We used TDD($\delta_{q_j}\boldsymbol{q}_j$) and TDD($\delta_{q_j}\boldsymbol{z}_{kj}$) distributions separately for each j and k to draw proposals for the allele frequencies. The δ_{q_j}'s are proposal parameters that are adjusted during the burn-in as in Ovaskainen et al. (2008) to give an accept ratio of 0.44.

We thus used the truncated Dirichlet distribution of Fang *et al*. (2000) to perform the Metropolis-Hastings random walk for the Dirichlet-distributed variables $\boldsymbol{\kappa}$, $\boldsymbol{q}$ and $\boldsymbol{z}$ with a pre-set truncation threshold $\tau = 10^{-7}$. This greatly improves the mixing properties of the Markov chain, because it helps to avoid numerical problems on the boundary of the parameter space (i.e. on the edges of the simplices Δ^{n_j-1} and $\Delta^{n_\mathcal{P}-1}$). According to our observation, the method that Fang *et al*. (2000) present for sampling from TDD may produce biased samples for high truncation thresholds such as $\tau = 10^{-1}$. However, to our experience, this does not compromise the statistical power of our algorithm with $\tau = 10^{-7}$.

We have implemented the algorithm described above in the R-package RAFM (Karhunen 2012).

References

Fang, K.T., Z. Geng and G.L. Tian, 2000 Statistical inference for the truncated Dirichlet distribution and its application in misclassification. Biometrical Journal **42**: 1053-1068.

Karhunen, M., 2012 RAFM: Admixture F-model. http://CRAN.R-project.org/package=RAFM.

Ovaskainen, O., H. Rekola, E. Meyke and E. Arjas, 2008 Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data. Ecology **89:** 542-554.