Supporting Text

Simulations. In general, the decrease of heterozygosity with time in a diploid population of size N due to genetic drift alone is approximately proportional to $e^{-t/(2N)}$ and linear in a relatively small range. We use simulations to predict the rate of decay of heterozygosity over time and/or space under specific evolutionary histories. Such simulations should accurately model the dynamics of observed data (Fig. 4*A*) if realistic values are used for all of the necessary demographic parameters; these values can be suggested from appropriate ethnographic/archaeological observations.

The model of a serial founder effect we used to generate Fig. 4*B* is that of a series of stepping stones during their formation process. This model is in contrast to the classical stepping stone model (1), which predicts genetic diversity once equilibrium has been reached between the evolutionary forces of mutation, migration, and drift. We therefore call our model the "dynamic stepping stone" (DSS), but there is no stable equilibrium: the curve connecting heterozygosity with geographic distance from the origin continues to change with time. When regressing expected heterozygosity on geographic distance from the origin, the tendency toward mutation–drift equilibrium will, slowly but eventually, result in a flat regression line through the whole distance range.

Under our model, a sample of size N_b from a large population of size K (the initial stepping stone) migrates to a new location where the sample founds a new colony that develops as an isolate and grows rapidly to the maximum population size/carrying capacity of K. The new colony can be seen as a newly formed stepping stone, and by the same process a group of founders of size N_b from it will give rise to a new colony further away. Sampling the last colony formed to generate a new stepping stone is repeated a total of n times. If the new colonies are formed at constant time intervals in a linear space at a constant distance from the previous colony in the chain, a row of n stepping stones arises at a constant rate in units of time t and space x. Considering only the drift effect of one bottleneck, one can then expect at the end of the expansion that the slope of the decay of expected heterozygosity is $-1/(2N_b)$ (2); its decay with distance from the origin x is $-x/(2N_b)$, or $-t/(2N_b)$ with time t since the origin, if x and t are estimated in terms of

numbers of bottlenecks. At the end of the expansion, the decay of heterozygosity from the first to the last colony will be given as a first approximation by Eq. **2**.

$\widetilde{\Delta H}$ differs from the simulated value (since $n = 100$ and $N_b = 250$, $\widetilde{\Delta H} = 0.2$ instead of 0.12) because Eq. **2** tends to underestimate the drift components listed above and ignores the counterbalancing effect of mutation. But Eq. **2** does show that the slope of the regression depends on the ratio n/N_b; if one wants to extract more biological information from the slope value, either n or N_b must be estimated from other sources.

Fig. 4*B* shows the results of a simulation of the process that might have produced Fig. 4*A*. The intercept of Fig. 4*B* differs somewhat from that of Fig. 4*A* and is affected by the model and parameters with which the mutation process is simulated. The mutation rate used in the simulation was 7.567×10^{-4}; this rate is the weighted average based on repeat motif [weighted by the number of di-, tri-, and tetranucleotide markers listed in Zhivotovsky *et al.* (3)] of the mutation rates used by Zhivotovsky *et al.* (3) for the Human Genome Diversity Project–Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) data set. We assumed that mutations were generated by a simple stepwise mutation model, with equal probability of increasing or decreasing a microsatellite allele by one repeat within a range of 0–25 repeats per locus, with hard reflecting boundaries at the endpoints of this range.

In the simulation, attention was extended to the components of the drift process ignored in Eq. **2**: the time between successive bottlenecks (T), the growth rate of the colonies, and the carrying capacity of the colonies (K). A generation time of 25 years was assumed (4). As discrete generations were used in the simulation, population sizes are "effective" ones, i.e., about one-third of census sizes of real populations (5).

The total time of expansion was assumed to be 50,000 years (2,000 generations). The size of the global population at the end of the expansion was $nK = 10^6$, of the order expected at the beginning of agriculture, after which the growth rate of populations increased. The creation of colonies occurred linearly in space. The basic assumption was that the founders of a new colony were one hunter–gatherer tribe, and therefore N_b was set at 250. There are historical examples of new colonies with the number of founders varying from a few dozen to a few thousand, and the census size of hunter–gatherer tribes varies in this range (4). To use realistic values of nT and nK and for the simplicity of

other calculations, n was set equal to 100, and T equal to 20, with the growth of the founders from N_b to K taking a little less than $T/2$ generations.

The simulation does not consider migration between the colonies, which would decrease the slope toward zero in the graph of Fig. 4*B*. Other factors that would tend to decrease the slope of the fitted line are stabilizing selection, which in the case of microsatellites would probably be due to their close linkage to loci with heterozygous advantage, or other complications like admixture with older inhabitants (6).

Most of the pitfalls of this model are avoided with another model, demic diffusion (DD) (), that uses a model of diffusion proposed for a related problem (8). This model may be more realistic because it depends on variables that are known with reasonable accuracy for hunter–gatherers (for example, camp size and migration and growth rates per generation); DD simulations have been used in earlier studies (5, 9–11). Under this model, a population is simulated as a set of niches forming a territory of any shape, although usually a rectangle, and each niche has a maximum carrying capacity. We also ran DD simulations to study the decay of expected heterozygosity with distance from the origin, using microsatellites. In our model, generations are discrete, and the population changes at every generation; every individual has a variable number of children who stay with probability $1 - m$ in the same niche of the parents and the rest move to any adjacent niche (north, south, east, or west) with equal probability. The variance of migration is m. The growth rate is the number of children per individual parent and must be greater than one for population growth to occur (a growth rate of 1.8 was used). If the number of individuals per niche in the new generation is greater than the carrying capacity, the surplus is culled randomly. Our DD simulations generate a linear decay of expected heterozygosity with distance, but we found it difficult to run them at realistic population sizes because DD is much more demanding of computer time than DSS.

1. Kimura, M. & Weiss, G. M. (1964) *Genetics* **49,** 561–576.
2. Hartl, D. L. & Clark, A. G. (1997) *Principles of Population Genetics* (Sinauer, Sunderland, MA), 3rd Ed., p. 172.
3. Zhivotovsky, L. A., Rosenberg, N. A. & Feldman, M. W. (2003) *Am. J. Hum. Genet.* **72,** 1171–1186.

4. Cavalli-Sforza, L. L., ed. (1986) *African Pygmies* (Academic, New York).
5. Cavalli-Sforza, L. L. & Bodmer, W. (1999) *The Genetics of Human Populations* (Dover, New York).
6. Eswaran, V., Harpending, H. & Rogers, A. R. (2005) *J. Hum. Evol.* **49,** 1–18.
7. Cavalli-Sforza, L. L. (2004) in *Examining the Farming/Language Dispersal Hypothesis*, eds. Bellwood, P. & Renfrew, C. (McDonald Institute Monographs, Cambridge, U.K.).
8. Fisher, R. A. (1937) *Ann. Eugenics* **7,** 355–369.
9. Sgaramella-Zonta, L. & Cavalli-Sforza, L. L. (1973) in *Genetic Structure of Populations*, ed. Morton, N. E. (Univ. Press of Hawaii, Honolulu), pp. 128–135.
10. Rendine, S., Piazza, A. & Cavalli-Sforza, L. L. (1986) *Am. Nat.* **128,** 681–706.
11. Edmonds, C. A., Lillie, A. S. & Cavalli-Sforza, L. L. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 975–979.