Mathematical constraints on F_{ST}: biallelic markers in arbitrarily many populations SUPPLEMENTARY FILE S1: rectangular and linear stepping-stone models

Nicolas Alcala[1] **and Noah A. Rosenberg**
Department of Biology, Stanford University, Stanford, CA 94305-5020, USA

In addition to the island migration model simulations in the main text, we also considered the rectangular and linear stepping-stone migration models. This supplementary text compares properties of M and F_{ST} observed in the three models.

Simulations

The simulation approach follows that used for the island model. In the rectangular stepping-stone model, subpopulations are arranged on a rectangular bounded habitat. Each subpopulation receives migrants from each adjacent subpopulation with the same probability. Subpopulations not on the habitat boundaries receive migrants at the same rate $\frac{m}{4}$ from each of four adjacent subpopulations; subpopulations on habitat edges receive migrants from each of three adjacent subpopulations at rate $\frac{m}{3}$; subpopulations at vertices receive migrants from each of two adjacent subpopulations at rate $\frac{m}{2}$ (Maruyama 1970).

In the linear stepping-stone model, subpopulations are arranged along a linear bounded habitat. Each subpopulation receives migrants from each adjacent subpopulation at the same rate: interior subpopulations receive migrants at rate $\frac{m}{2}$ from each of two adjacent subpopulations, whereas subpopulations at habitat boundaries receive migrants from a single adjacent subpopulation at rate m (Maruyama 1970).

Figure S1 provides a schematic of the parametrization of all three models. For $K = 2$, all three migration models (island, rectangular stepping-stone, linear stepping-stone) are equivalent. Under the rectangular stepping-stone model, for $K = 7$, we considered a 4×2 habitat with one subpopulation missing at the edge; for $K = 40$, we considered an 8×5 habitat.

Results

Rectangular stepping-stone model. Under the rectangular stepping-stone model, properties of F_{ST} in relation to M are qualitatively similar to those under the island model, but with higher F_{ST} (Figures S2B,E,H and S3B,E,H). For a fixed number of subpopulations K, the geometry in the rectangular stepping-stone model, with 2 to 4 connections per subpopulation, generates less migration among the subpopulations, so that the genetic difference among subpopulations is higher than in the fully connected graph of the island model. Thus, with M, K, and $4Nm$ held constant, F_{ST} is generally higher in the rectangular stepping-stone model.

Linear stepping-stone model. In the linear stepping-stone model, F_{ST} is higher still than in the rectangular model (Figures S2C,F,I and S3C,F,I). Connectivity among subpopulations is reduced, with each subpopulation having only 1 or 2 neighbors. The probability that a mutation remains localized and fixed in some subpopulations while being absent in others is greater than in the other geometries, so that F_{ST} exceeds that observed in the other models.

Proximity of the joint density of M and F to the upper bound in stepping-stone models. For a fixed number of subpopulations and a fixed scaled migration rate, $\bar{F}_{ST}/\bar{F}_{\max}$ is smaller under the island model than under the rectangular stepping-stone model, and smaller under the rectangular model than under the linear model (Figure S4). This observation can be explained by the stronger constraints on migration in the linear case, in which immigrants come from at most 2 other subpopulations, than in the rectangular case, with up to 4 neighbors, and the island model, with $K - 1$. The smaller number of neighbors prevents genetic homogenization between subpopulations and thus leads to larger F_{ST} values.

Compared to the relatively limited effect on $\bar{F}_{ST}/\bar{F}_{\max}$ in the island model of the number of subpopulations K (Figure S4A), under the rectangular and linear stepping-stone models, K has a stronger influence on $\bar{F}_{ST}/\bar{F}_{\max}$—which increases with K (Figure S4B,C). This result can be explained by noting that unlike in the island model, which is fully connected irrespective of the number of subpopulations, at a fixed migration rate, the increasing number of subpopulations produces greater isolation of distant subpopulations in the stepping-stone models, generating greater genetic differentiation and thus leading to F_{ST} values closer to their upper bounds. Our results thus provide a more precise formulation of the classical pattern that under the linear stepping-stone model, F_{ST} values tend to be closer to 1 than under an island model (Maruyama 1970): under stepping-stone models, F_{ST} values tend to be closer to their upper bound in terms of M than under an island model.

Literature Cited

Maruyama, T., 1970 Effective number of alleles in a subdivided population. Theor. Pop. Biol. **1**: 273–306.

Manuscript compiled: Friday 5th May, 2017%
[1]Correspondence to Nicolas Alcala. Email: nalcala@stanford.edu.

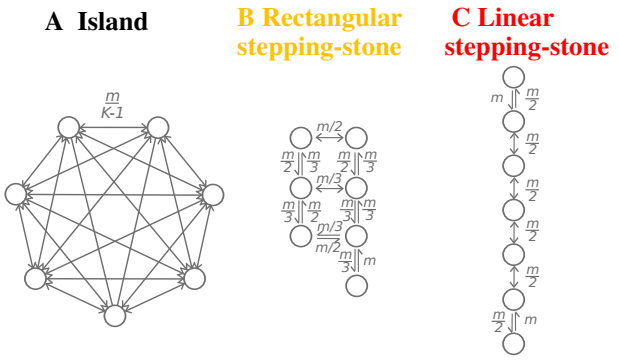

Figure S1 Three migration models. (A) Island model. (B) Rectangular stepping-stone model. (C) Linear stepping-stone model. Quantities on the arrows represent the backward migration rates between pairs of subpopulations.

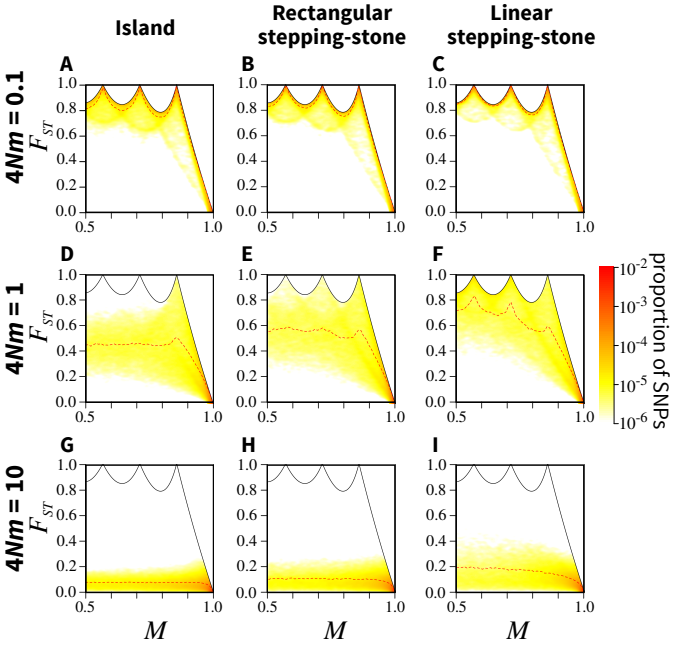

Figure S2 Joint density of the frequency M of the most frequent allele and F_{ST}, for different migration models and scaled migration rates $4Nm$, considering $K = 7$ subpopulations. Panels A,D,G for the island model are copied from Figure 3B,E,H for ease of comparison. The simulation procedure and figure design follow Figure 3.

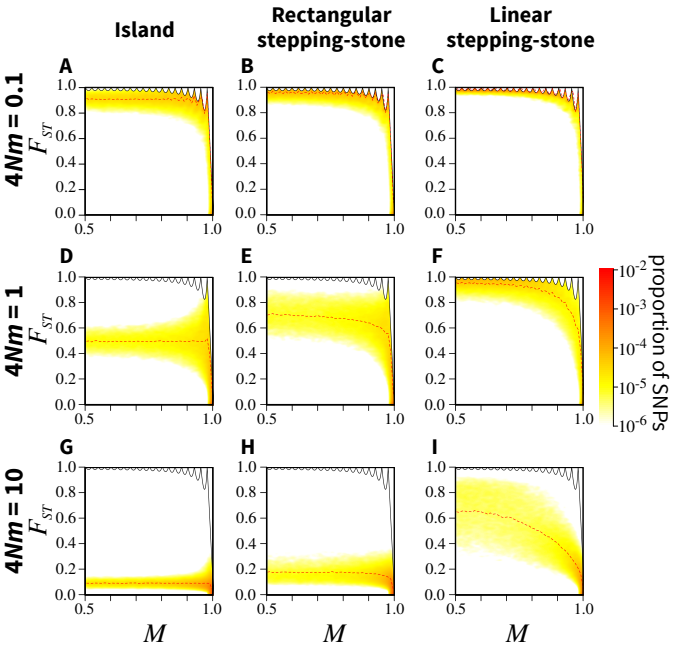

Figure S3 Joint density of the frequency M of the most frequent allele and F_{ST}, for different migration models and scaled migration rates $4Nm$, considering $K = 40$ subpopulations. Panels A,D,G for the island model are copied from Figure 3C,F,I for ease of comparison. The simulation procedure and figure design follow Figure 3.

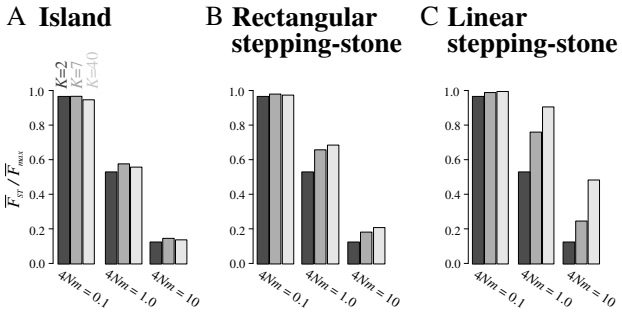

Figure S4 $\bar{F}_{ST}/\bar{F}_{\max}$, the ratio of the the mean F_{ST} to the mean maximal F_{ST} given the observed frequency M of the most frequent allele, as a function of the number of subpopulations K and the scaled migration rate $4Nm$, for three migration models. (a) Island model. (b) Rectangular stepping-stone model. (c) Linear stepping-stone model. Panel A for the island model is copied from Figure 4 for ease of comparison. The simulation procedure and figure design follow Figure 4.

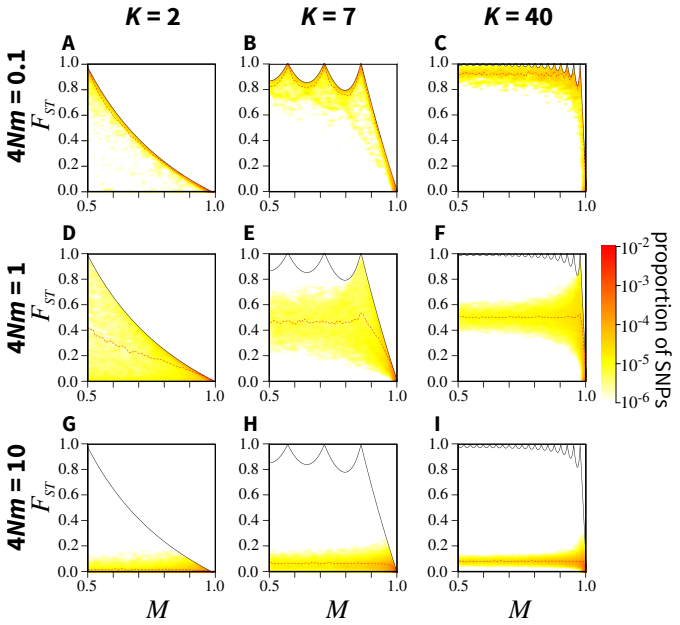

Figure S5 Joint density of the frequency M of the most frequent allele and F_{ST}, for different number of subpopulations K and scaled migration rates $4Nm$, with SNPs simulated under a infinitely-many-sites model. (A) $K = 2, 4Nm = 0.1$. (B) $K = 7, 4Nm = 0.1$. (C) $K = 40, 4Nm = 0.1$. (D) $K = 2, 4Nm = 1$. (E) $K = 7, 4Nm = 1$. (F) $K = 40, 4Nm = 1$. (G) $K = 2, 4Nm = 10$. (H) $K = 7, 4Nm = 10$. (I) $K = 40, 4Nm = 10$. SNPs are simulated using coalescent software `MS`, assuming an island model of migration, a scaled mutation rate $\theta = 1$, and 100,000 replicate simulations, with 100 lineages sampled per subpopulation. For $K = 2$, $\theta = 1/\sum_{i=1}^{199} \frac{1}{i} \approx 0.1702698$. For $K = 7$, $\theta = 1/\sum_{i=1}^{699} \frac{1}{i} \approx 0.1403$. For $K = 40$, $\theta = 1/\sum_{i=1}^{3999} \frac{1}{i} \approx 0.1127251$. To obtain single-SNP simulations, simulations with more than one segregating sites were discarded. The figure design follows Figure 3.