# DISPERSAL, GENE FLOW, AND ISOLATION

MODELS OF POPULATION STRUCTURE AND GENE FLOW

One-Way Migration and Genetic Admixture

Gene flow between a large continental population and a small island can be treated as a one-way migration. Although dispersal occurs in both directions, the genetic impact of migrants on the larger continental population is small and can generally be ignored.

Imagine an allele with different frequencies in the continent (Q) and island (q_0) populations. The proportion of the island population comprising migrants in any generation is defined as m, and the proportion comprising natives is $(1 - m)$.

Continent

Q

m

Island

q_0

$1 - m$
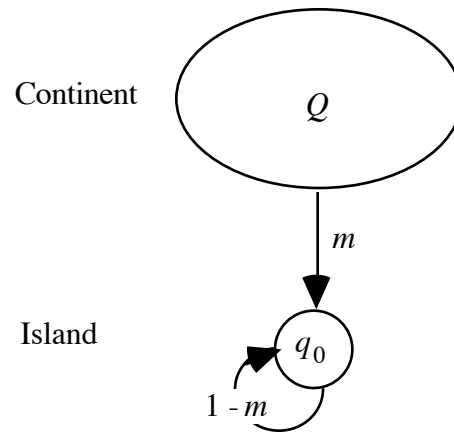
After one generation of migration, the frequency of the allele on the island is

$$q_1 = (1 - m)q_0 + mQ = q_0 - mq_0 + mQ = q_0 - m(q_0 - Q).$$

The change in allelic frequency due to migration into the island population is

$$\Delta q = q_1 - q_0 = [q_0 - m(q_0 - Q)] - q_0 = -m(q_0 - Q).$$

Solving for m, $$m = \frac{q_1 - q_0}{Q - q_0} .$$ The coefficient m represents the proportion of genes from the migrant source in one generation.

The **admixture coefficient** (M), or the cumulative effect of m for t generations is $$M_t = \frac{q_t - q_0}{Q - q_0} .$$

Let $1 - m = \dfrac{q_1 - Q}{q_0 - Q},$ Glass and Li (1953)showed that if one-way gene flow persists for t generations at the rate of m per generation.

Convince yourself this equation is true by deriving it from the earlier equation or *m*.

Thus, $(1 - m)^t = \dfrac{q_t - Q}{q_0 - Q}$, which can be rearranged as $m \approx 1 - e^{\ln(x)/t}$, where $x = \dfrac{q_t - Q}{q_0 - Q}$.

Glass and Li used this model to determine the rate of Caucasian gene flow (m) into the African-American population living in the region of Baltimore, MD. They assumed one-way gene flow in their model (although it is obviously 2-way); because the products of African ´ Caucasian matings are culturally defined as African-American in North America. They solved for m by assuming t was about 10 generations (roughly 300 years).

Some of the data used by Glass and Li are listed below:

Allele		West Africans q_0	American blacks q_t	American whites Q	M	$(1 - m)^t$	m
Rh	r	0.22	0.27	0.38	0.313	0.687	0.037
PTC	t	0.18	0.30	0.55	0.324	0.676	0.038
ABO	B	0.17	0.13	0.08	0.444	0.556	0.057
average across many alleles					~0.300		0.035

Would you expect variance in estimates of m from different alleles? Why?

Averaged across many alleles, M was about 0.30 and m about 0.035 per generation. Glass and Li concluded that North American blacks in Baltimore obtained about 30% of their genes from admixture with Caucasians and 70% from their West African ancestors.

A weakness of this analysis is that they could not specify which West African populations were ancestral to N. American blacks. Considerable variance in gene frequencies occurs among African populations.

Adams and Ward (1973) reviewed studies of admixture in other populations. For example, m for American blacks from Oakland, CA, was about 0.020; and from Claxton, GA, was about 0.009 per generation. Why do you think these regional differences in admixture exist?

WRIGHT'S ISLAND MODEL

Sewell Wright (1931) imagined an ancestral population that was broken up into an array of partially isolated demes, each of effective size N. In the absence of selection, genetic differentiation among subpopulations results from an equilibrium between genetic drift (a function of N) and gene flow (m). At equilibrium, and for small values of m, F_{ST} is related to gene flow in the following way:

$$F_{ST} \approx \frac{1}{4Nm + 1}$$
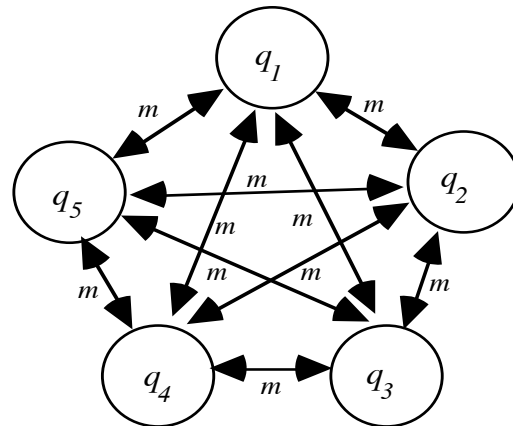
Nm can be interpreted as the effective <u>number</u> of migrants exchanged between demes per generation (Wright 1969). As you can see from the graph, as *Nm* increases, F_{ST} decreases, and *vice versa*.

Note: that $Nm \gg 1$ results in little divergence (i.e., low F_{ST}). However, low migration (e.g., $Nm = 0.25$) results in a large $F_{ST} = 0.50$.
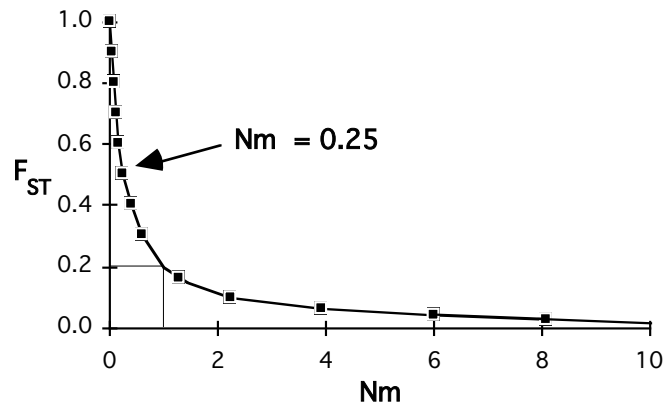

In general, Nm < 1 results in relatively large values for F_{ST}. Wright interpreted *Nm* as a virtual number of migrants—i.e., the virtual number of migrant individuals it would take to produce the observed F_{ST} at equilibrium between drift and gene flow. This can be useful, as we shall soon see.
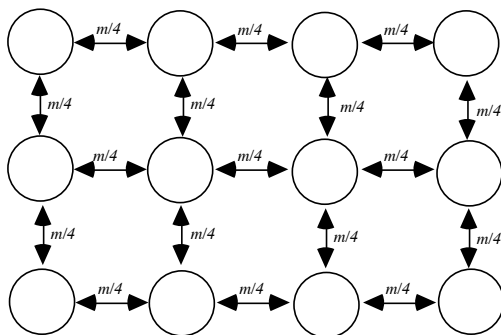
Stepping-Stone Models

For organisms with relatively low effective rates of dispersal, most gene exchange will occur between neighboring demes.

Linear Stepping-Stone
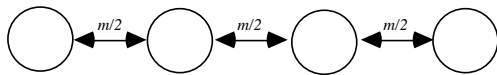
This pattern of migration is common for species distributed in linear habitats (e.g., estuaries along a coast).

Two-dimensional Stepping-Stone

Oases in a desert might be distributed as a two-dimensional stepping-stone. Kimura and Weiss (1964) showed that the correlation in gene frequencies (*r*) between demes decreases approximately exponentially as a function of the number of steps (*x*) between demes.

For a 1-dimensional stepping stone, $r \propto e^{-x\sqrt{2\mu/m}}$. For two dimensions, $r \propto (e^{-x\sqrt{4\mu/m}})\sqrt{x}$, where *m* is short-distance migration rate, μ is long-distance migration rate (essentially the same as the mutation rate), and $m \gg \mu$.. Below., I have plotted these relationships, assuming that μ was roughly equivalent to the mutation rate, or about 10^{-6}, and $m = 0.001$.

Note: the correlation in allelic frequencies (*r*) between pairs of populations falls off more steeply with distance as more dimensions are added to the stepping stone model.
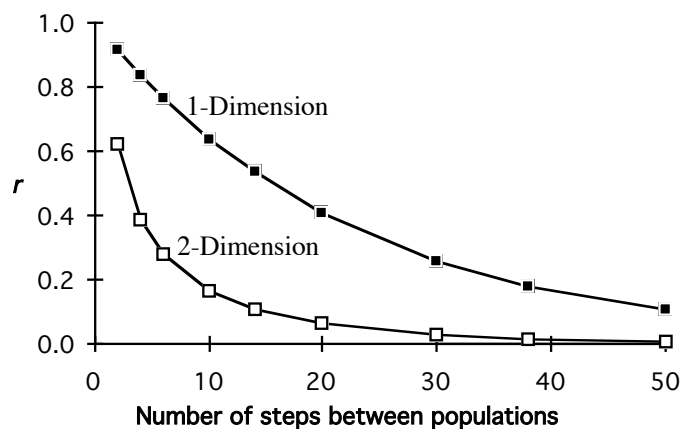
Also **note**: the genetic correlation between demes stays high longer with higher rates of migration.

For example, the adjacent 1-dimensional stepping stone shows different values of *m*.. Lond distance migration, μ, was held constant at 10^{-6}. Obviously the slopes of these negative correlations provide information about dispersal rates.
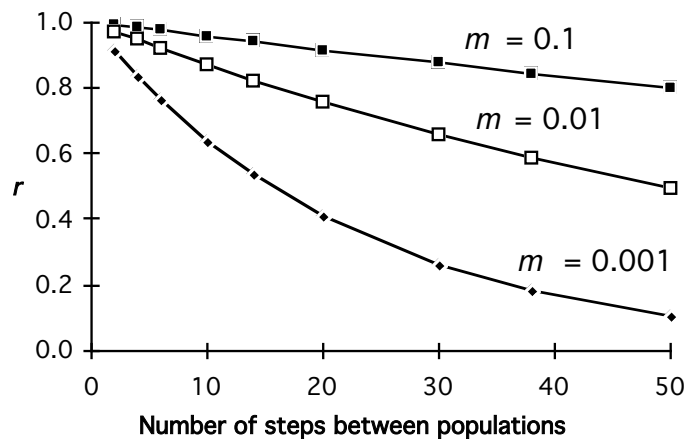
Isolation-by-Distance

Isolation-by-distance in a continuously distributed population produces a result that is resembles stepping-stone models. The genetic correlation between individuals declines as a function of geographic distance between them. The math is complicated, but you should refer to following published treatments for details (Malécot 1955; Wright 1943).
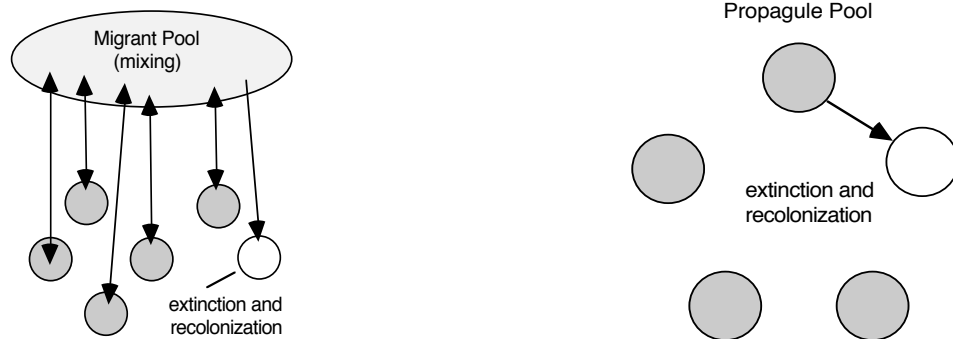
Metapopulation Models

If local extinctions and recolonization events occur frequently, we need to consider metapopulation models (Gilpin 1991; Levins 1970), because divergence among demes also affected by extinction and recolonization rates.

Wright (1940; 1977) suggested that the founder events associated with recolonization should increase the differentiation (i.e., F_{ST}) among subpopulations.

Slatkin (1977) investigated this issue with two variations of Wright's (1931) basic island model. In the **migrant-pool** model, *k* colonists are drawn from a random sample of the entire metapopulation. In the **propagule-pool** model, *k* colonists derive from a single

population. These different patterns of recolonization will determine whether F_{ST} will be greater or less than the expectations of the basic island model.

Extinction and recolonization events add sampling variance to the metapopulation. Recolonization increases the rate of genetic drift if founder population sizes tend to be small. The source of colonists (migrant-pool vs. propagule-pool) also affects F_{ST}. Differentiation may be slowed down if colonists are drawn from neighboring propagules.

Maruyama and Kimura (1980) explored propagule-pool colonization under island and stepping-stone models. With high extinction and recolonization rates, the effective size of the metapopulation is less than nN (n is the number of demes and N the average deme size). This reduction in N_e results in the retention of less genetic diversity (H_t) for the metapopulation as a whole.

Wade and McCauley (1988) used computer simulations to examine the effects of migrant pool and propagule pool models on differentiation. They found that F_{ST} is primarily affected by several factors:

- the mode of colonization of empty patches (migrant-pool vs. propagule-pool)
- k, the average number of colonists
- m, the average rate of dispersal between subpopulations
- N, local population sizes

The extinction rate (e) seems to have very little effect on F_{ST}.

Metapopulation coalescence

Gilpin (1991) described the adjacent example of **coalescence** of subpopulation genotypes (a, b, c) in a metapopulation with a high rate of extinction. Local extinction events (horizontal bars) and recolonization events (dashed arrows) occur in three discrete subpopulations (1, 2, 3). Note that in time all subpopulations end up with genotype c. This process graphically illustrates the Maruyama/Kimura effect—i.e., populations become uniform among and homozygous within.
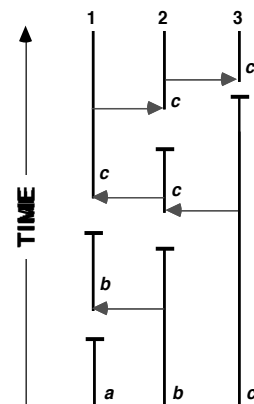
ESTIMATION OF GENE FLOW

F_{ST}–methods

Several indirect methods exist for estimating gene flow from gene frequency data (Slatkin 1985a; Slatkin & Barton 1989). The F_{ST}–method is presently the most widely used approach.

Because we can estimate F_{ST} empirically from gene frequency data, Nm can be inferred from the following relationship:

$$Nm \approx \left(\frac{1 - F_{ST}}{4F_{ST}} \right)$$

Convince yourself that this relationship is true by deriving this equation from an earlier statement about the relationship between F_{ST} and Nm..

To make this inference, we assume the following conditions are met:

- selective neutrality of the genetic markers
- equilibrium has been reached between migration and genetic drift
- an island-model of distribution

Based on simulations that violated each of these assumptions, Slatkin and Barton (1989) found that estimates of Nm based on the F_{ST}–methods are robust compared to other methods. For example:

Species	F_{ST}	Nm
major human racial groups	0.07	3.30
subpopulations of Ord's desert rat	0.70	0.11
riverine subpopulations of the Sonoran topminnow	0.50	0.25

Isolation-by-distance

Slatkin (1993) considered estimation of migration rates under stepping-stone and isolation-by-distance models with equilibrium and non-equilibrium conditions. Under both models, estimated migration rates between pairs of populations should decline with increasing distance between populations.

Slatkin developed a computer program (DIST) that calculates multilocus F_{ST}'s for all pairs of demes. Both weighted (Weir & Cockerham 1984) and unweighted (Nei 1973) estimates of F_{ST} can be used. The pairwise F_{ST}'s are used to estimate:

$$\hat{M} \approx \frac{1 - G_{ST}}{4G_{ST}},$$

where $\hat{M}$ is an estimator of Nm for a pair of populations.

A handy version of the DIST program can be found in GENEPOP 3.1b (Raymond & Rousset 1995) email:Raymond@isem.univ-montp2.fr

Regression of the $\hat{M}$ values against geographic distance (G) allows detection of an isolation-by-distance affect.

$\hat{M}$ declines linearly as a function of distance. The slope of this decline is a function of the dimensionality of the stepping-stone model. A slope of 1.0 is expected under a linear stepping-stone model and a slope of 0.5 is expected under a 2-dimensional stepping-stone model.
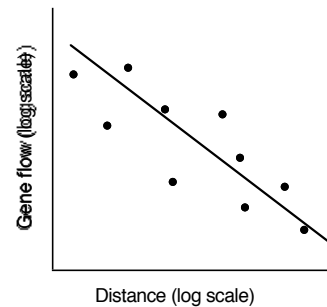
This relationship was found in the deep-sea hydrothermal vent tube worm, *Riftia pachyptila* (Black *et al.* 1994). Gene flow declined linearly with increasing distance between sites distributed along the East Pacific Rise, a linear archipelago of vent habitats spread along a mid-ocean spread-ing center. This finding suggests that most dispersal if tubeworm larvae occurs be-tween neighboring populations.
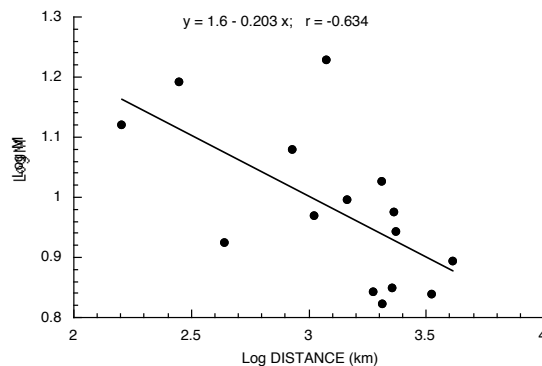

The significance of this relationship can be detected with Mantel's permutation test. A parametric correlation analysis is not appropriate because the distance values are not in-dependent of one another--they are pair-wise distances. The Mantel procedure is best de-scribed in (Manly 1991). A nice package of programs that allow you to perform a Mantel test is called the **R Package** (Legendre 1991):

http://www.fas.umontreal.ca/BIOL/Casgrain/R/

The null hypothesis (Island model)

The null hypothesis, of course, is that there is no relationship between gene flow and geographical distance. This would be expected under Wright's basic island model where migrants are drawn from a random pool that represents all source popula-tions. Species with long-distance dispersal abilities are expected to show this pattern.

We found this pattern in the deep-sea hydrothermal vent mussel *Bathymodiolus thermophilus* (Craddock *et al.* 1995). Mussels have a feeding larval stage that is capable of long-distance dispersal in the water column; they are thoroughly mixed between sites,
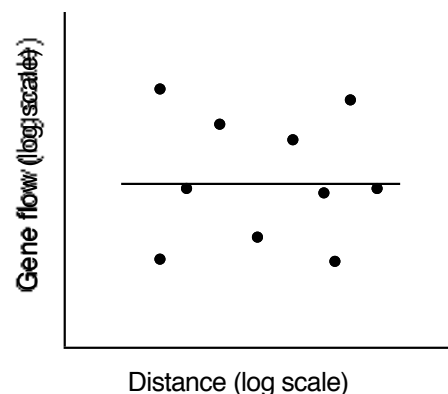
Both mtDNA and allozymes were used and they produced the same basic result. Note, however, that the appropriate estimator for gene flow rate based on mitochondrial variation is

$$\hat{M} \approx \frac{1 - G_{ST}}{2G_{ST}}$$

What has changed and why should this be right?

Non-equilibrium conditions

The F_{ST}–method can also be used to identify recent stepwise range expansions. It can also be used on a variety of population models (e.g., island model, stepping stone, etc.). However, we interpret the *Nm* estimates as if they had come from an idealized island model population structure at equilibrium. Estimates of *Nm* in a stepping-stone system will be underestimates. Why?

Other Methods

Private Alleles

The "private alleles" technique provides an alternative approach to estimating Nm (Slatkin 1985b); however, this technique requires a sufficient number of private alleles (alleles that occur in only one deme) and reasonably large sample sizes to obtain unbiased estimates. The method is based on the idea that under high gene flow, alleles that occur in only one deme will be at very low frequency because they are probably new mutations that have not yet had time to spread.

The F_{ST} method is less sensitive to sample size if a sufficient number of polymorphic loci are examined (Nei 1973). In an empirical analysis of near-shore marine fishes Waples (1987) found that the F_{ST} approach provided a better fit to expectations of dispersal rates than did the private alleles method. Slatkin and Barton (1989) came to the same conclusion. Different estimators of F_{ST} (e.g., Nei 1973; Weir & Cockerham 1984) perform about equally well for moderate to low levels of gene flow.

Genealogical Methods

More on this in the PowerPoint presentation given during lecture. Genealogical methods also exist for estimating *Nm* and isolation by distance from dis-crete molecular states (e.g., Beerli 1998; Neigel 1997; Slatkin & Maddison 1989, 1990). These methods require the construction of gene tress. The methods are based on the premise that each unique sequence arises once by mutation in one place. If more than one population has a particular sequence, it must have migrated to the other places. We will get back to these approaches when we cover genealogical methods later in the course.

REFERENCES

Adams, J, Ward, RH (1973) Admixture studies and the detection of selection. *Science,* **180,** 1137-1143.

Beerli, P (1998) Estimation of migration rates and population sizes in geographically structured populations. In: *Advances in Molecular Ecology* (ed. Carvalho G), pp. (in press). ISO, Amsterdam.

Black, MB, Lutz, RA, Vrijenhoek, RC (1994) Gene flow among vestimentiferan tube worm (*Riftia pachyptila*) populations from hydrothermal vents of the Eastern Pacific. *Marine Biology,* **120,** 33-39.

Craddock, C, Hoeh, WR, Lutz, RA, Vrijenhoek, RC (1995) Extensive gene flow in the deep-sea hydrothermal vent mytilid *Bathymodiolus thermophilus*. *Marine Biology,* **124,** 137-146.

Gilpin, ME (1991) The genetic effective size of a metapopulation. *Biological Journal of the Linnaean Society,* **42,** 165-175.

Glass, B, Li, CC (1953) The dynamics of racial admixture—an analysis based on the American Negro. *American Journal of Human Genetics,* **5,** 1-20.

Kimura, M, Weiss, WH (1964) The stepping stone model of genetic structure and the decrease of genetic correlation with distance. *Genetics,* **49,** 561-576.

Legendre, P (1991) *The R Package: Multidimensional analysis, spatial analysis*, 3rd edn. Département de sciences biologiques, Université de Montréal, Montréal, Quebec.

Levins, R (1970) Extinction. In: *Some Mathematical Questions in Biology* (ed. Gerstenhaber M), pp. 75-108. American Mathematical Society, Providence, R.I.

Malécot, G (1955) Remarks on decrease of relationship with distance, following paper by M. Kimura. *Cold Spring Harbor Symp. Quant Biol.,* **20,** 52-53.

Manly, BFJ (1991) *Randomization and Monte Carlo Methods in Biology*. Chapmann and Hall, New York.

Maruyama, T, Kimura, M (1980) Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Sciences, USA,* **77,** 6710-6714.

Nei, M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA,* **70,** 3321-3323.

Neigel, JE (1997) A comparison of alternative strategies for estimating gene fow from genetic markers. *Annual Review of Ecology and Systematics,* **28,** 105-128.

Raymond, M, Rousset, F (1995) GENEPOP: (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity,* **86,** 248-249.

Slatkin, M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology,* **12,** 253-262.

Slatkin, M (1985a) Gene flow in natural populations. *Annual Review of Ecology and Systematics,* **16,** 393-430.

Slatkin, M (1985b) Rare alleles as indicators of gene flow. *Evolution,* **39,** 53-65.

Slatkin, M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution,* **47,** 264-279.

Slatkin, M, Barton, NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution, 43,* 1349-1368.

Slatkin, M, Maddison, WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics, 123,* 603-613.

Slatkin, M, Maddison, WP (1990) Detecting isolation by distance using phylogenies of genes. *Genetics, 126,* 249-260.

Wade, MJ, McCauley, DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution, 42,* 995-1005.

Waples, RS (1987) A multispecies approach to the analysis of gene flow in marine shore fishes. *Evolution, 41,* 385-400.

Weir, BS, Cockerham, CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution, ´38,* 1358-1370.

Wright, S (1931) Evolution in Mendelian populations. *Genetics, 16,* 97-159.

Wright, S (1940) Breeding structure of populations in relation to speciation. *American Naturalist, 74,* 232-248.

Wright, S (1943) Isolation by distance. *Genetics, 28,* 114-138.

Wright, S (1969) *Evolution and the Genetics of Populations: The Theory of Gene Frequencies.* The University of Chicago Press, Chicago, Illinois.

Wright, S (1977) *Evolution and the Genetics of Populations: Experimental Results and Evolutionary Deductions.* The University of Chicago Press, Chicago, Illinois.