

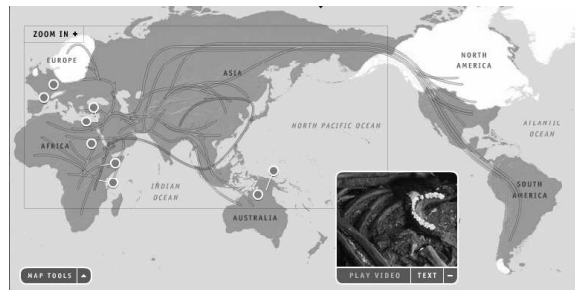
Contents

1 Migration	1
1.1 Introduction	1
1.2 Allele Migration Model	1
1.3 Genotype Migration Model	2
1.4 Whalund Effect	2
1.5 Linkage Disequilibrium & Migration	3
1.6 Estimating Migration	4
1.7 Other Models	4
1.7.1 One-Island Model	4
1.7.2 Island Model	5
1.7.3 Stepping-Stone Model	6
1.7.4 General Model	6

1 Migration

1.1 Introduction

Migration Introduction



Migration is the movement of individuals between populations. Until now we have considered a single, panmictic population. (Panmictic means the population is completely mixed; there are no subpopulations or substructures within the population.)

Now, we consider a population with immigration and emmigration.

1.2 Allele Migration Model

Allele Migration Model

Assume that who migrates is determined independent of genotype.

Assume two populations: the one we study and another one from which migrants arrive. Assume both of those populations are panmictic.

Then, let m be the fraction of the population we study in each generation that are new immigrants. We assume m is not dependent on time, so each generation immigrants make up the same proportion of the population.

We need to track allele frequencies in two populations. Suppose the allele frequency of a gene of interest in our study population is p_1 and in the foreign population is p_2 .

The recursion relation for the allele frequency in our study population is

$$p_1(t+1) = (1-m)p_1(t) + mp_2(t).$$

Extension to Multiple Alleles or Populations

The equation

$$p_1(t+1) = (1-m)p_1(t) + mp_2(t).$$

applies to any number of alleles. All alleles satisfy the same equation. We need only determine the allele's frequency in both populations.

If we allow for n distinct populations, with population 1 our study population, then we need to define n m_i migration proportions, where m_i is the proportion of the study population that came from population i . Then

$$p_1(t+1) = \left(1 - \sum_{i=2}^n m_i\right) p_1(t) + \sum_{i=2}^n m_i p_i(t).$$

Allele Frequency Changes

The change in allele frequency in one generation due to migration from one other population is

$$\Delta p_1(t+1) = p_1(t+1) - p_1(t) = m[p_2(t) - p_1(t)]$$

so there is a change which depends on the difference in allele frequencies between the two populations. We must make further assumptions to determine what will happen in the long run (see 1.7), but now we can conclude there will be continual change in the allele frequency as long as there are immigrants coming from populations with different allele frequencies.

1.3 Genotype Migration Model

Migration at the Genotype Level

In diploid populations, the immigrants are diploid and carry two genes in with them.

Let $P_{ij}^{(k)}(t)$ be the frequency of genotype ij in population k at generation t . Then, for two populations

$$P_{ij}^{(1)}(t+1) = (1-m)P_{ij}^{(1)}(t) + mP_{ij}^{(2)}(t).$$

What does migration do to Hardy Weinberg equilibrium? It can disrupt it. To see this, consider the example where $P_{AA}^{(1)}(t) = 1$ and $P_{BB}^{(2)}(t) = 1$. Then, after one generation of migration,

$$\begin{aligned} P_{AA}^{(1)}(t+1) &= (1-m) \\ P_{BB}^{(1)}(t+1) &= m. \end{aligned}$$

Is this population in Hardy Weinberg equilibrium?

1.4 Whalund Effect

Whalund Effect

In general, for multiple populations each with allele A frequency p_i , we have

$$P_{AA}^{(1)}(t+1) = \sum_i m_i P_{AA}^{(i)}(t),$$

where we write $m_1 = 1 - \sum_{i=2}^n m_i$.

If the populations are each at HWE, then

$$P_{AA}^{(1)}(t+1) = \sum_i m_i p_i^2(t).$$

Now, examine the above equation. The right side is just the definition of the expectation of a discrete random variable p , where $P(p = p_i) = m_i$. So, we can write

$$P_{AA}^{(1)}(t+1) = E(p^2),$$

where the expectation is taken over populations.

Now, recall the study population has allele frequency after migration of

$$p_1(t+1) = \left(1 - \sum_{i=2}^n m_i\right) p_1(t) + \sum_{i=2}^n m_i p_i(t) = \sum_{i=1}^n m_i p_i(t),$$

but this is recognizable as

$$p_1(t+1) = E(p).$$

If the study population were in HWE, then

$$P_{AA}^{(1)}(t+1) = [E(p)]^2,$$

but it isn't and the deviation from HWE is

$$P_{AA}^{(1)}(t+1) - [E(p)]^2 = E(p^2) - [E(p)]^2 = \text{Var}[p(t)].$$

We conclude then, that

$$P_{AA}^{(1)}(t+1) = [E(p)]^2 + \text{Var}[p(t)].$$

In words, the frequency of homozygotes in a population receiving immigrants from multiple population is increased over Hardy-Weinberg expected frequencies by a factor equalling the variance in allele frequencies across different populations and weighted by their relative contributions.

This is yet another reason why we may not observe HWE in a sample from a population. If instead of sampling from one population, we are actually sampling from multiple populations in proportions m_i , then the presence of homozygotes will be increased above expectation.

In the two allele case, heterozygotes will be deficient. In the multi-allele case, some heterozygotes may be more common than expected, but overall there will be fewer heterozygotes than expected under HWE.

Life of Whalund Effect

How persistent is the Whalund effect? It is true immediately after migration, but what happens if migration stops and this population is allowed to randomly mate?

What happens if sperm and eggs are the objects that are migrating, *not* adults?

1.5 Linkage Disequilibrium & Migration

Linkage Disequilibrium and Migration

We have already seen an example of mixing two populations in linkage equilibrium that would create huge linkage disequilibrium. Consider a population of *AACC* individuals and another with *BBDD* individuals. Then, after equal mixing the population produces only *AC* and *BD* gametes and now *AD* or *BC* gametes though they should be equally likely.

To consider the case in greater detail, let p_i be the frequency of allele *A* at locus one in population i and q_i be the frequency of allele *C* at locus two in population i .

Consider an individual producing gametes. If the individual immigrated from population i , and population i is in linkage equilibrium, then this individual will produce gamete *AC* with probability $p_i q_i$. Overall,

$$P_{AC}^{(1)}(t+1) = \sum_{i=1}^n m_i p_i q_i.$$

Linkage Disequilibrium and Migration

The allele frequencies in the study population are as usual

$$\begin{aligned}p_A^{(1)}(t+1) &= E(p) \\ p_C^{(1)}(t+1) &= E(q).\end{aligned}$$

The amount of linkage disequilibrium is

$$\begin{aligned}D_{AC}^{(1)}(t+1) &= \sum_{i=1}^n m_i p_i q_i - E(p) E(q) \\ &= E(pq) - E(p) E(q) \\ &= \text{Cov}(p, q).\end{aligned}$$

So the amount of linkage disequilibrium in the study population is given by the covariance in the allele frequencies over populations, weighted by the relative contributions of those populations to the study population.

Gene Flow

Migration can introduce a very temporary Hardy Weinberg Disequilibrium that disappears after one generation of random mating.

Migration can introduce linkage disequilibrium that can disappear slowly with random mating.

Eventually, a new Hardy Weinberg and linkage equilibrium will be achieved. The only long-term impact of migration is to change the allele frequencies in the study population. The new allele frequencies are the average of allele frequencies in the populations.

Since migration in the long-run only changes allele frequencies, migration is often referred to as *gene flow* and the populations that are receiving or contributing alleles are called *gene pools*.

1.6 Estimating Migration

Estimating Migration

If the three populations are identified, the two sources of immigration and the mixture population that results, and the allele frequencies in each population are known, then the rate of migration is simple to estimate using the recurrence relation

$$p_1(t+1) = (1-m)p_1(t) + mp_2(t).$$

Solve for m to obtain

$$m = \frac{p_1(t+1) - p_1(t)}{p_2(t) - p_1(t)}$$

Exercise. How would you get a variance on this one?

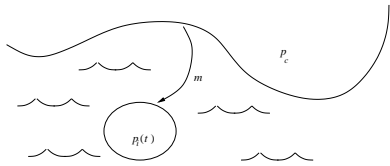
Application: Estimating flow of a biogenic gene used in a crop plant to surrounding wild plant populations.

1.7 Other Models

We did not discuss these other models in detail, but I include them for completeness. You should understand how to set up models like these and the main conclusions, e.g. what happens in the long-run. Details are less important because these models are pretty specific and limited in applicability.

1.7.1 One-Island Model

One-Island Model



m is the migration rate from the continent, p_c is the fixed allele frequency on the continent, and $p_i(t)$ is the variable allele frequency on the island.

In the one-island model, there is a small island neighboring a large continent with different allele frequencies. Migration from island to continent is too small to affect the continent, but migration from continent to island changes allele frequency on the island. The recurrence relation for allele frequency on the island is

$$p_i(t + 1) = (1 - m)p_i(t) + mp_c,$$

One-Island Model Equilibrium

We would like to determine the equilibrium gene frequency on the island $p_i(t) \rightarrow p_i$. Solve the equation by dropping dependence on t ,

$$p_i = (1 - m)p_i + mp_c$$

yields $p_i = p_c$. So, eventually the island allele frequency will match the continent allele frequency.

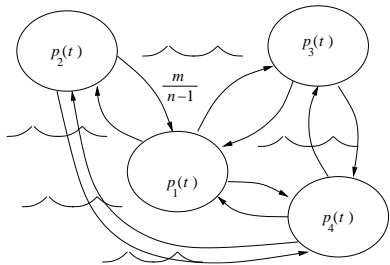
The rate of approach to equilibrium is determined by m . To see, consider

$$\begin{aligned} p_i(t + 1) - p_c &= (1 - m)p_i(t) - p_c + mp_c \\ &= (1 - m)(p_i(t) - p_c). \end{aligned}$$

So, each generation the amount of disequilibrium $p_i(t) - p_c$ declines by fraction m . The more migration, the faster the island matches the continent.

1.7.2 Island Model

Island Model



$\frac{m}{n-1}$ is the migration rate between islands and $p_i(t)$ is the variable allele frequency on the i th island.

Assume there are n islands. In each generation, every island provides a fraction $\frac{m}{n-1}$ to every other island, i.e. a fraction m of the population is in flux every generation, and there is completely balanced exchange of individuals between islands. Population i allele frequency is updated as

$$p_i(t + 1) = (1 - m)p_i(t) + \sum_{j \neq i} \frac{m}{n-1} p_j(t).$$

Island Model Equilibrium

At equilibrium, the allele frequency in population i is p_i and the following equation is satisfied.

$$\begin{aligned}
 p_i &= (1 - m)p_i + \sum_{j \neq i} \frac{m}{n-1} p_j \\
 &= \left(1 - m - \frac{m}{n-1}\right) p_i + \frac{m}{n-1} \sum_j p_j \\
 &= \left(1 - \frac{mn}{n-1}\right) p_i + \frac{mn}{n-1} \bar{p} \\
 p_i &= \bar{p}.
 \end{aligned}$$

In addition, one can show $\bar{p}(t) = \frac{1}{n} \sum_i p_i(t)$ is unchanging with respect to time.

So, the equilibrium allele frequency that will settle on every population is the average allele frequency across all islands. Migration does not change the overall allele frequency, just its distribution around the islands.

Island Model Rate of Approach to Equil.

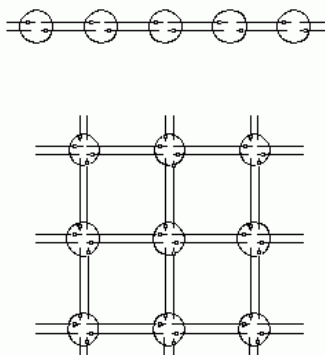
The rate of approach is

$$\begin{aligned}
 p_i(t+1) - \bar{p} &= \left(1 - \frac{mn}{n-1}\right) p_i(t) + \frac{mn}{n-1} \bar{p}(t) - \bar{p} \\
 &= \left(1 - \frac{mn}{n-1}\right) (p_i(t) - \bar{p}),
 \end{aligned}$$

because $\bar{p}(t) = \bar{p}$. So the rate of approach is like that of the one-island model with $m^* = \frac{mn}{n-1}$. Notice, that for n large, $\frac{mn}{n-1} \approx m$, so the models are essentially equivalent when there are lots of islands.

1.7.3 Stepping-Stone Model

Stepping-Stone Model



In this model migration only occurs between neighbors. For the linear model $\frac{m}{2}$ of the population in the next generation comes from each neighbor. For the 2-dim model $\frac{m}{4}$ of the population in the next generation comes from each neighbor.

More complex models of migration exist, for example, where the migration rate $m(d)$ depends on the distance d between populations.

1.7.4 General Model

General Migration Matrix

Assume n populations and let m_{ij} be the fraction of individuals in population i that are new immigrants from population j at each generation. Then, we can write

$$p(t+1) = Mp(t),$$

where $p(t) = (p_1, p_2, \dots, p_n)^t$ is the vector of allele frequencies in the n populations and superscript t indicates transpose.

Equivalently, we can let m_{ij}^* be the fraction of population i individuals that emigrate to population j each generation. Then,

$$p(t+1)^t = p(t)^t M^*.$$

Summary of Migration

- All populations achieve the same equilibrium allele frequency if at every generation each population retains some of its population (not everyone emigrates) and all populations can receive immigrants from all other populations, even if it takes multiple generations. (Mathematically: If the Markov chain M is irreducible and ergodic, then there exists a limiting distribution independent of the starting state, and since all alleles achieve the same equilibrium distribution, the allele frequencies will be the same across islands.)
- The rate of approach to equilibrium is controlled by m .

In general, migration eliminates geographic differences and the speed with which it does this is determined by the amount of migration m .