Migration-Selection models in Population Genetics

Abhilesh Dhawanjewar

April, 28, 2017

Contents

Introduction

Geography plays a central role in the pattern of genetic differentiation in a species. Geographically isolated populations will accumulate mutations over time increasing the differentiation between them. Migration between populations acts to counter this process through genetic exchange and leading to homogenization of the populations. Using population genetic models I aim to investigate the dynamics of allele frequencies on a focal island population under different migratory scenarios.

Continent-Island Model

The model

- Unidirectional migration from a large mainland population, the *continent* to a smaller population, the *island* (Fig: 1).
- The frequency of an allele A on a biallelic genetic locus is p_c on the continent and p_t on the island in generation t
- Let m be the proportion of alleles on the island that come from the continent in each generation, also referred to as the migration rate or the rate of gene flow.
- After one generation, for the island population, a proportion m would have come from the continent and a proportion *(1 - m)* would come from the island.
- The equation for the allele frequency of A would therefore be -

$$p_{t+1} = (m)p_c + (1-m)p_t = p_t + m(p_c - p_t)$$

$$\implies \Delta p = p_{t+1} - p_t = m(p_c - p_t)$$

- At equilibrium, we set $\Delta p = 0$ and get $\hat{p} = p_c$
- Performing this recursively, $p_t = p_c + (1-m)^t(p_0 - p_c)$, where p_0 is the initial allele frequency on the island.

The code

```
# Defining the default intial parameters
p_c = 0.6  # Allele frequency on the continent
p_0 = 0.1  # Initial allele frequency on the island
m = 0.01  # rate of gene flow
n_gen = 500  # number of generations
```
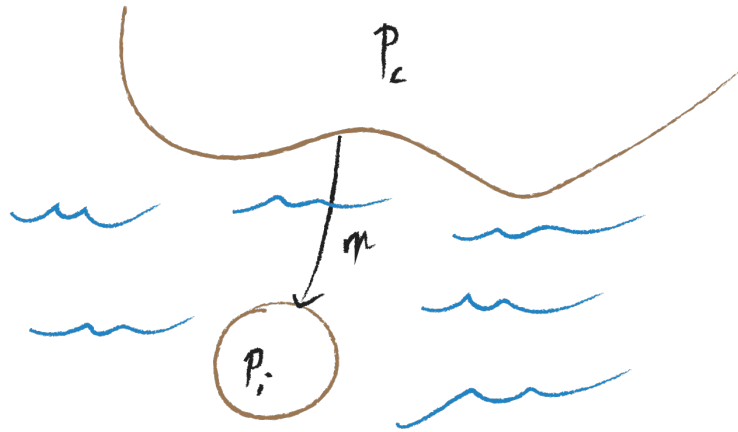
Figure 1: **Continent-island model.** m is the rate of migration from the continent with an allele frequency p_c to the island with an allele frequency p_t

```r
counter = 1

# Define the color palette for the plots
col_pal = rep(c("gray0", "firebrick3", "dodgerblue3"), 2)

sim_func <- function(p_c, p_0, m, n_gen) {

    p_t <- c(1:n_gen)
    p_t[1] = p_0

    for (i in 2:n_gen) {
        p_t[i] = p_c + (1 - m)^i * (p_0 - p_c)
    }

    if (counter %in% c(1, 4)) {
        plot(p_t, type = "l", xlab = "Time in generations", ylab = "Allele frequency",
            ylim = c(0:1), cex.lab = 0.7, cex.axis = 0.7, mgp = c(2,
                1, 0), col = col_pal[[counter]])
    } else {
        points(p_t, type = "l", col = col_pal[[counter]])
    }
}

layout(matrix(c(1, 2), nrow = 1, ncol = 2, byrow = T))

# Run the simulation for different values of migration rate
vect_m = c(0.1, 0.01, 0.005)

for (j in vect_m) {
    sim_func(p_c, p_0, j, n_gen)
    counter <- counter + 1
}

legend("topright", legend = c("m = 0.1", "m = 0.01", "m = 0.005"),
    col = col_pal[1:3], lty = 1, cex = 0.6, box.lty = 0)
```
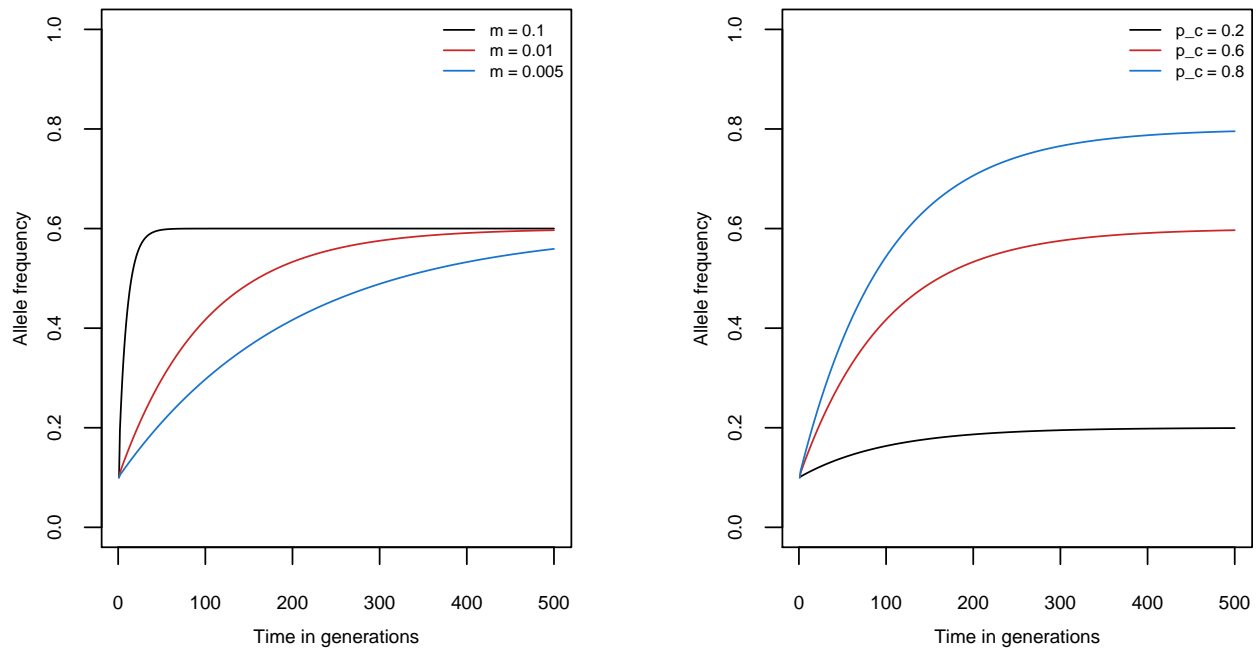
Figure 2: Effect of migration rate and continent allele frequency on the allele frequency on the island

```
# Run the simulation for different values of continent allele
# frequencies.
vect_p = c(0.2, 0.6, 0.8)

for (k in vect_p) {
    sim_func(k, p_0, m, n_gen)
    counter <- counter + 1
}

legend("topright", legend = c("p_c = 0.2", "p_c = 0.6", "p_c = 0.8"),
    col = col_pal[1:3], lty = 1, cex = 0.6, box.lty = 0)
```

Inferences

In the case of unidirectional migration from a large continent to a smaller island, the allele frequency on the island at convergence is the same as the allele frequency on the continent. The rate at which equilibrium is reached depends on the rate of migration from the continent to the island, with equilibrium being reached faster for higher rates of gene flow.
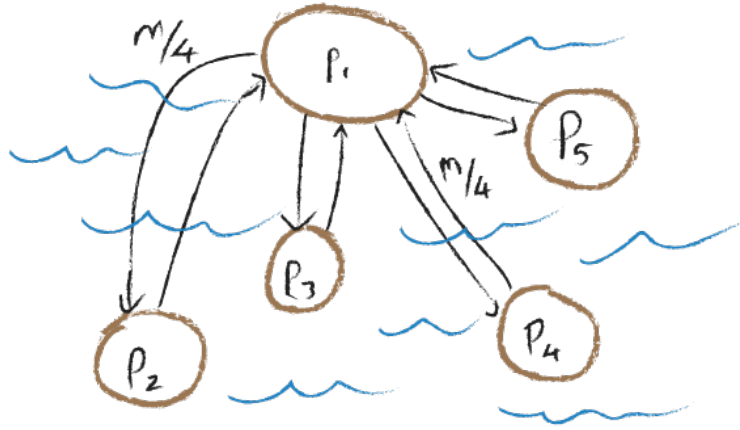
Figure 3: **Island-Island model.** Islands with allele frequencies p_1, p_2, p_3, p_4, p_5 exchange migrants at the rate of m per generation. The change in allele frequency on island 1 is investigated.

Island-Island model

The model

- A large population is geographically split into n subpopulations (islands) (Fig: 3).
- Each island consists of an equal proportion of m migrants from each other.
- The islands have allele frequencies $p_1, p_2, p_3, ..., p_n$
- Then the allele frequency on the island *1* after a generation as a result from migration from island *2* only will be -

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

- Generalizing to account for migrants arriving from all the islands, we get -

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

- Performing this recursively, we get -

$$p_t = t \sum_{i=2}^{n} m_i p_i (1 - \sum_{i=2}^{n} m_i) + p_0(1 - t\sum_{i=2}^{n} m_i + \sum_{i=2}^{n} m_i^t)$$

The code

```
# In this model I will assume there are five islands Defining
# the default initial parameters
p_0 = 0
p_1 = 0.6
p_2 = 0.6
p_3 = 0.6
p_4 = 0.6
m = 0.01
n_gen = 500


counter = 1


col_pal = rep(c("gray0", "firebrick3", "dodgerblue3"), 2)


sim_func <- function(p_0, p_1, p_2, p_3, p_4, m, n_gen) {
```

```r
    # Initial the simulation vector
    p_t <- c(1:n_gen)
    p_t[1] = p_0
    sum_migrants = m * p_1 + m * p_2 + m * p_3 + m * p_4

    # Run
    for (i in 2:n_gen) {
        p_t[i] = sum_migrants + (1 - 4 * m) * p_t[i - 1]
    }

    if (counter %in% c(1, 4)) {
        plot(p_t, type = "l", xlab = "Time in generations", ylab = "Allele frequency",
            ylim = c(0:1), cex.lab = 0.7, cex.axis = 0.7, mgp = c(2,
                1, 0), col = col_pal[[counter]])
    } else {
        points(p_t, type = "l", col = col_pal[[counter]])
    }
}

layout(matrix(c(1, 2), nrow = 1, ncol = 2, byrow = T))

vect_m = c(0.1, 0.01, 0.005)

for (i in vect_m) {
    do.call(sim_func, list(p_0, p_1, p_2, p_3, p_4, i, n_gen))
    counter = counter + 1
}

legend("topright", legend = c("m = 0.1", "m = 0.01", "m = 0.005"),
    col = col_pal[1:3], lty = 1, cex = 0.6, box.lty = 0)

vect_p = c(0.3, 0.4, 0.8)

for (j in vect_p) {
    do.call(sim_func, list(p_0, j, j, j, j, m, n_gen))
    counter = counter + 1
}

legend("topright", legend = c("avg(p) = 0.2", "avg(p) = 0.6",
    "avg(p) = 0.8"), col = col_pal[1:3], lty = 1, cex = 0.6,
    box.lty = 0)
```

Inferences

The island-island model looks at the change in allele frequency of an allele A on a particular island (say, island X) with migrants arriving from all the other island populations. In a simple case, where each island provides an equal number of migrants, the equilibrium allele frequency on island X is equal to the mean of the allele frequencies on the remaining islands. The rate of approach to equilibrium is dependent on the rate of migration per generation (m), the higher the rate of migration, the faster equilibrium is reached on island X. This is equivalent to treating the rest of the islands together as a continent population.
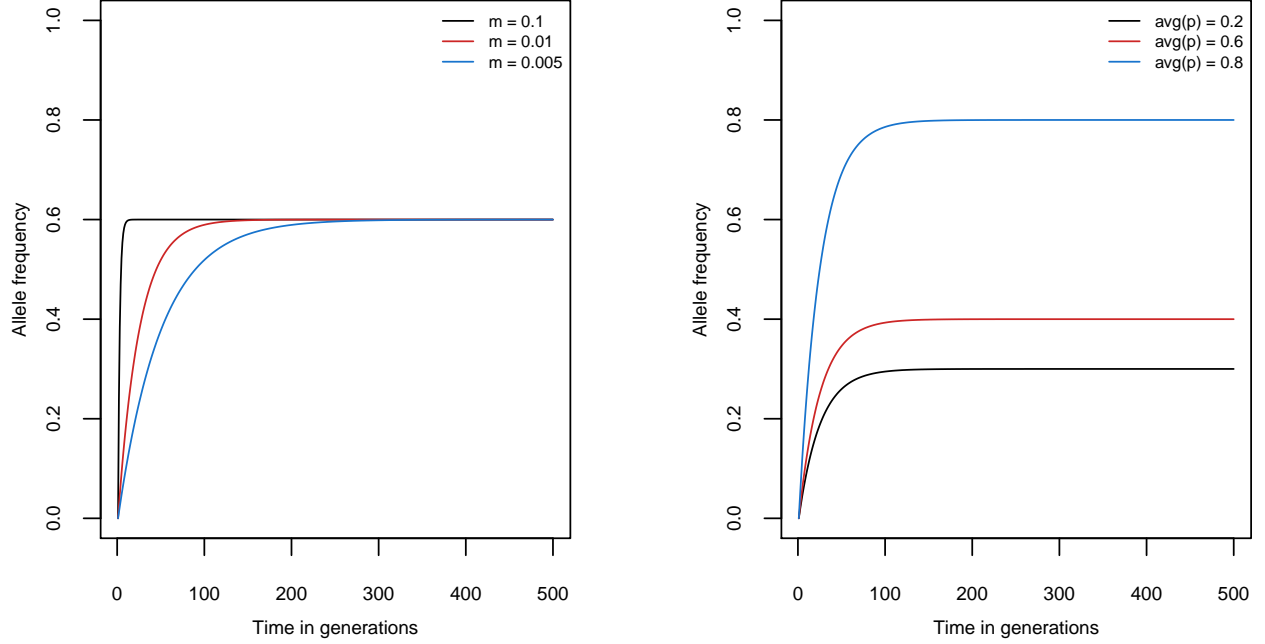
Figure 4: Effect of migration rate and mean allele frequency on all islands on the allele frequency on the island
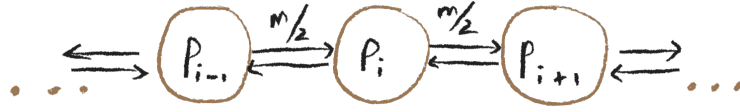

Figure 5: **One-Dimensional Stepping Stone model.** Infinite islands are arranged on a one-dimensional array. Each island exchanges $\frac{m}{2}$ migrants from its adjacent islands per generation.

One-Dimensional Stepping Stone Model

The model

- The population consists of an infinite one-dimensional array of n colonies (Fig: 5)
- In each generation, an individual can migrate at most "one-step" in either direction between the colonies. Exchange of migrants is restricted to between adjacent colonies.
- Let p_i denote the relative frequency of the allele A on the i th colony.
- Each island contributes $m_1/2$ migrants to its neighboring islands.
- This model also includes a long-range dispersal term and m_∞ stands for the rate of long range dispersal per generation i.e. the rate at which a colony exchanges individuals in each generation with a sample taken from the entire population in which the frequency of the allele A is $\bar{p}$
- The change in allele frequency on the i th colony after a generation will be -

$$p_i' = \frac{m_1}{2}(p_{i-1} + p_{i+1}) + (1 - m_i - m_\infty)p_i + m_\infty\bar{p}$$

- Generalizing this over multiple generations -

$$p_{(i,t+1)} = (1 - \sum_{i=1}^{\infty} m_i - m_\infty)p_{(i,t)} + \frac{m_i}{2}(p_{i-2} + p_{i-1} + p_i + p_{i+1} + p_{i+2}) + m_\infty\bar{p}(\frac{m_1}{2} + 2)$$

However, I could not analytically derive the recursive relation to calculate the allele frequency on the focal island after t generations.

6

Future Work

- Derive the recursive relationship for the one-dimensional stepping stone model
- Extend the stepping stone model to higher dimensions
- Extend the models to investigate the dynamics of multiple alleles
- Extend the models to incorporate alleles with different effective population sizes (Autosomes, Sex Chromosomes, Mitochondrial.)