

# Stochastic population genetics: homework 3

## To be returned on May 2

April 25, 2019

### 1 Coalescence with migration

In this problem we are interested in two populations of  $N \gg 1$  diploids living on two separate "islands". At each generation, parent alleles are picked at random (as in a W-F model) from the same island with probability  $1 - M/4N$  and from the other island with probability  $M/4N$  ( $M = O(1)$ ). Mutations occur independently of parents origin at rate  $\theta/4N$ . All mutations are assumed to be neutral and we take the limit of infinite size genome (i.e mutations never occur twice at the same site and always form a unique new genotype).

As in the W-F model, we consider each allele of each individual independently effectively giving two subpopulations of  $2N$  individuals. We note  $P_c(\theta)$  the probability that two alleles are the same if they come from the same subpopulation and  $P_d(\theta)$  if they come from two different subpopulations.

a. Why can we write that at first order the probability for two alleles from two different islands to have the same parent is 0? Show that

$$P_c(\theta) = \frac{1}{1 + \theta + M} + \frac{M}{1 + \theta + M} P_d(\theta). \quad (1)$$

b. Show that

$$P_d(\theta) = \frac{M}{1 + \theta + M} P_c(\theta). \quad (2)$$

Compute  $P_c(\theta)$  and  $P_d(\theta)$  as functions of the parameters.

c. How would the result of the previous question change if we had  $q$  islands and that the destination of each migration was chosen at random with equal probability within the islands?

d. Let  $t_c$  be the time to the first common ancestor between two individuals from the same island and  $t_d$  that same time for two individuals from different islands. Using a probabilistic argument, show that

$$P_c(\theta) = \int_0^\infty du e^{-u\theta} P(t_c = u), \quad (3)$$

and equivalently

$$P_d(\theta) = \int_0^\infty du e^{-u\theta} P(t_d = u). \quad (4)$$

e. From Eq. 3 and Eq. 4 show that

$$-P'_d(0) = \langle t_d \rangle \text{ and that } -P'_c(0) = \langle t_c \rangle. \quad (5)$$

Compute these two expectation values for  $q$  islands. What is the expected number of differences between two alleles? Why do you think both the average time to common ancestor and the average number of differences between two alleles of the same subpopulation are independent of the migration rate?