

# Population Structure and Gene Flow

---

COMP 571 - Fall 2010  
Luay Nakhleh, Rice University

# Outline

- (1) Genetic populations
- (2) Direct measures of gene flow
- (3) Fixation indices
- (4) Population subdivision and the Wahlund effect
- (5) Models of population structure
- (6) Population structure and the coalescent

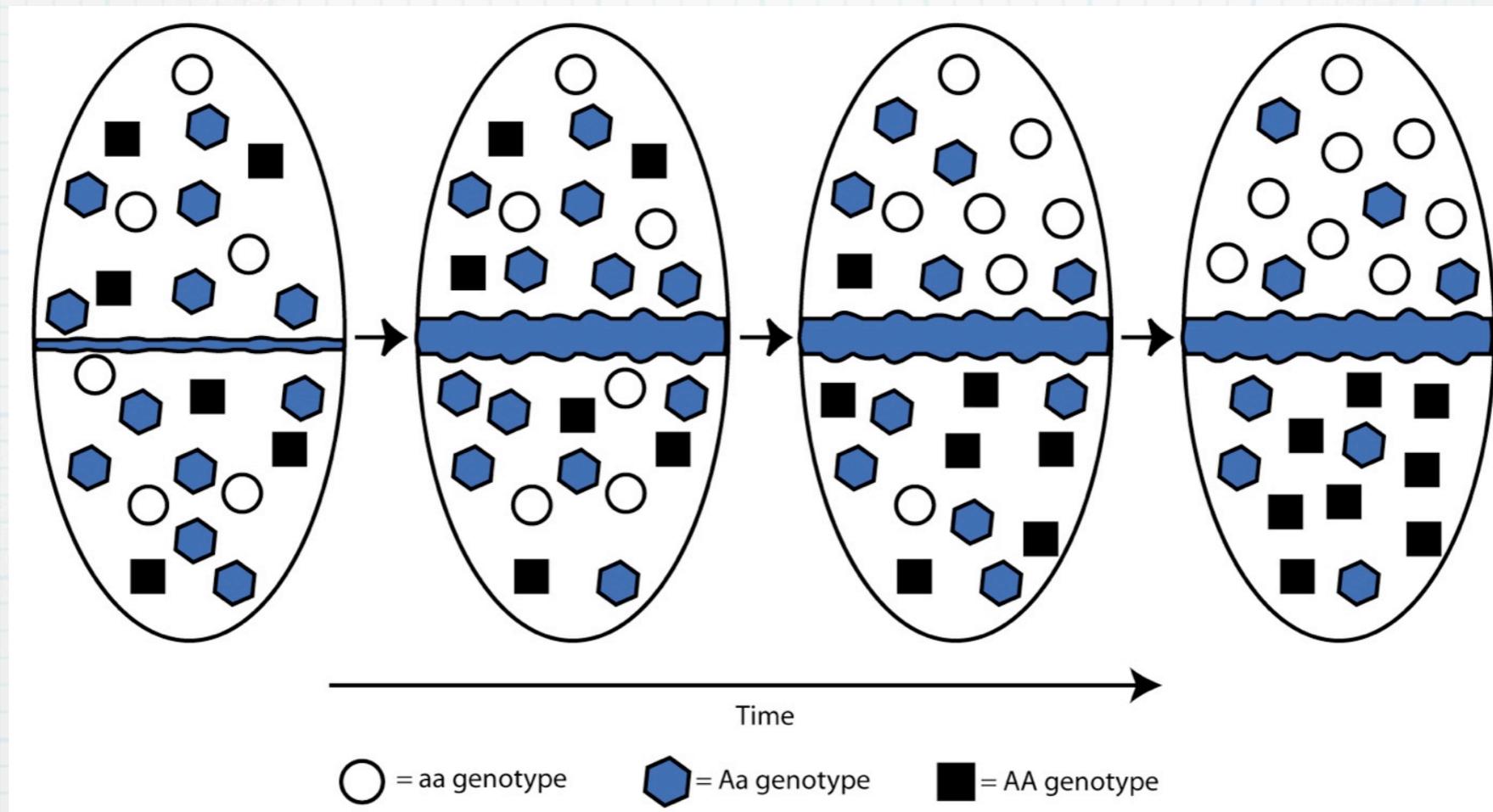
# (1) Genetic Populations

- \* Random mating is one of the assumptions underlying HW
- \* Implicit is the view that a population is a single entity where processes such as mating and movement of individuals are uniform throughout, a condition often called **panmixia**
- \* This is often not the case: chance that two individuals mate often depends on their location within the population

# (1) Genetic Populations

- \* **Population structure:** heterogeneity across a population in the chances that two randomly chosen individuals will mate
- \* **Gene flow:** the successful movement of alleles into populations through the movement of individuals (migration) or the movement of gametes

# (1) Genetic Populations



An example of population structure and allele-frequency divergence produced by limited gene flow.

A **Vicariance** event: the appearance of a geographic barrier (the river in the figure) that restricts gene flow among populations.

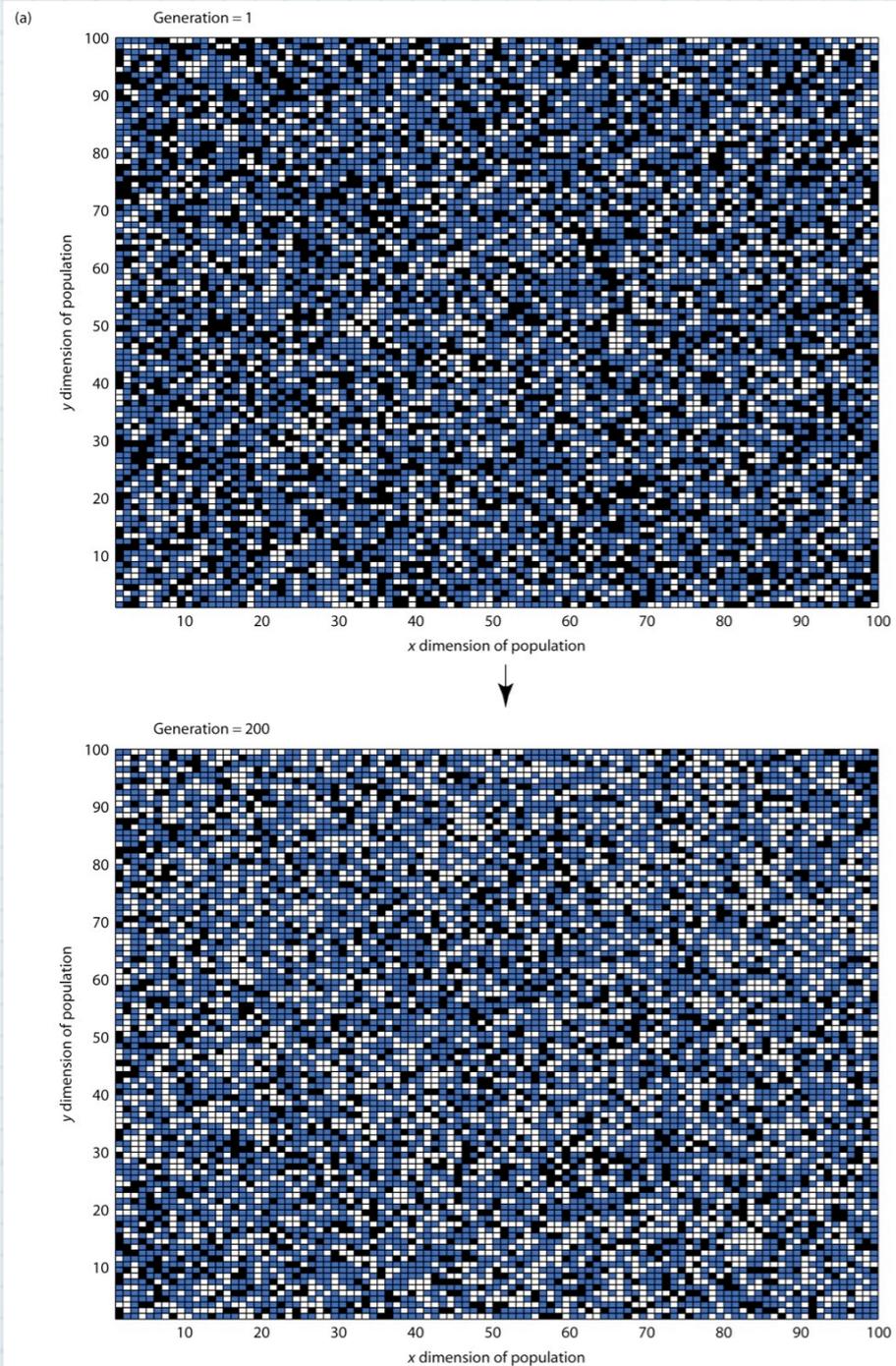
# (1) Genetic Populations



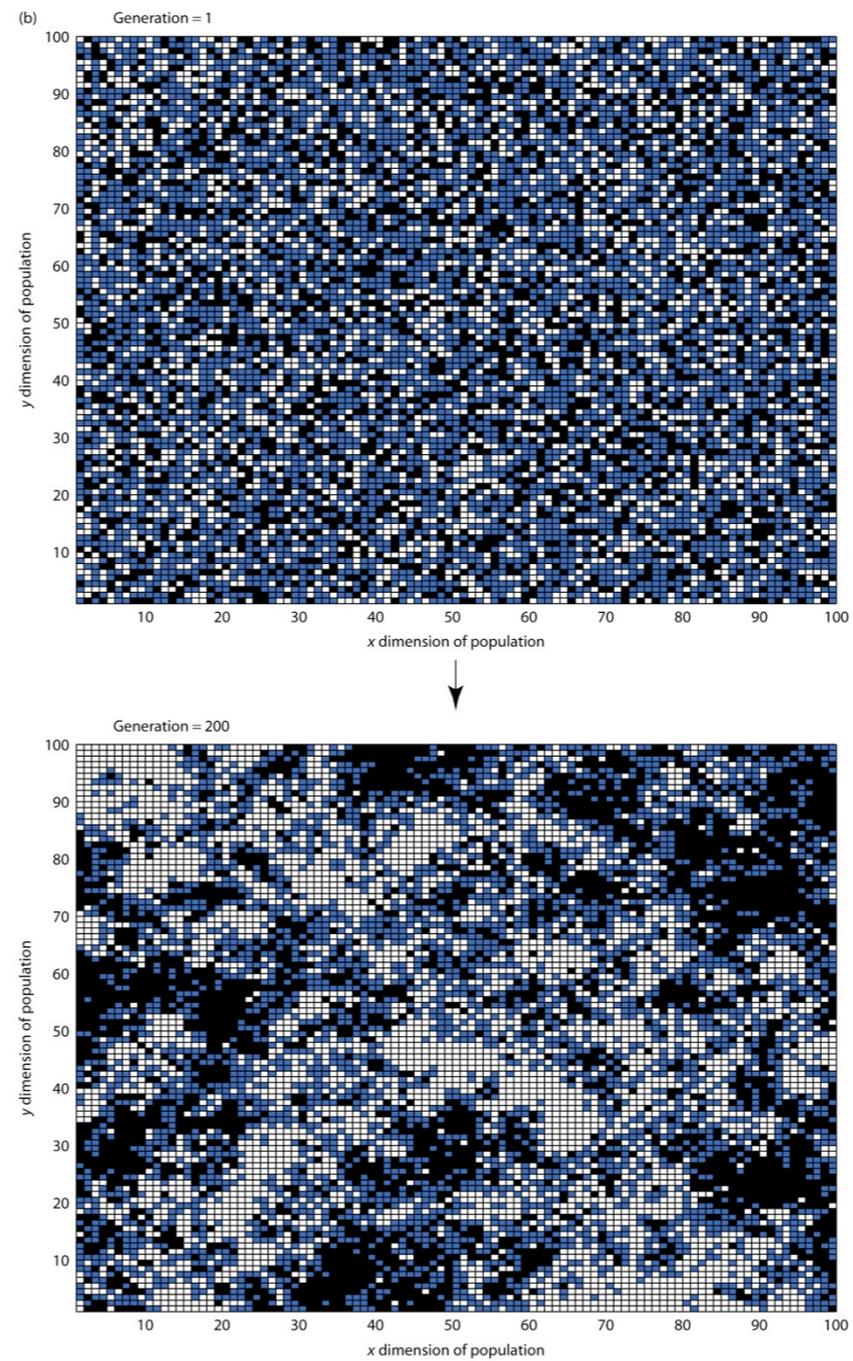
*L. parryae* has either blue or white flowers. In some locations most plants have blue or most plants have white flowers, whereas in other locations more equal frequencies of the two flower colors are found.

**Isolation by distance:** the decreasing chances of mating with increasing distance separating individuals.

# (1) Genetic Populations



Random mating over the entire population



The mating neighborhood is  $3 \times 3$  squares

■ heterozygous  
□ homozygous  
■ homozygous

# (1) Genetic Populations

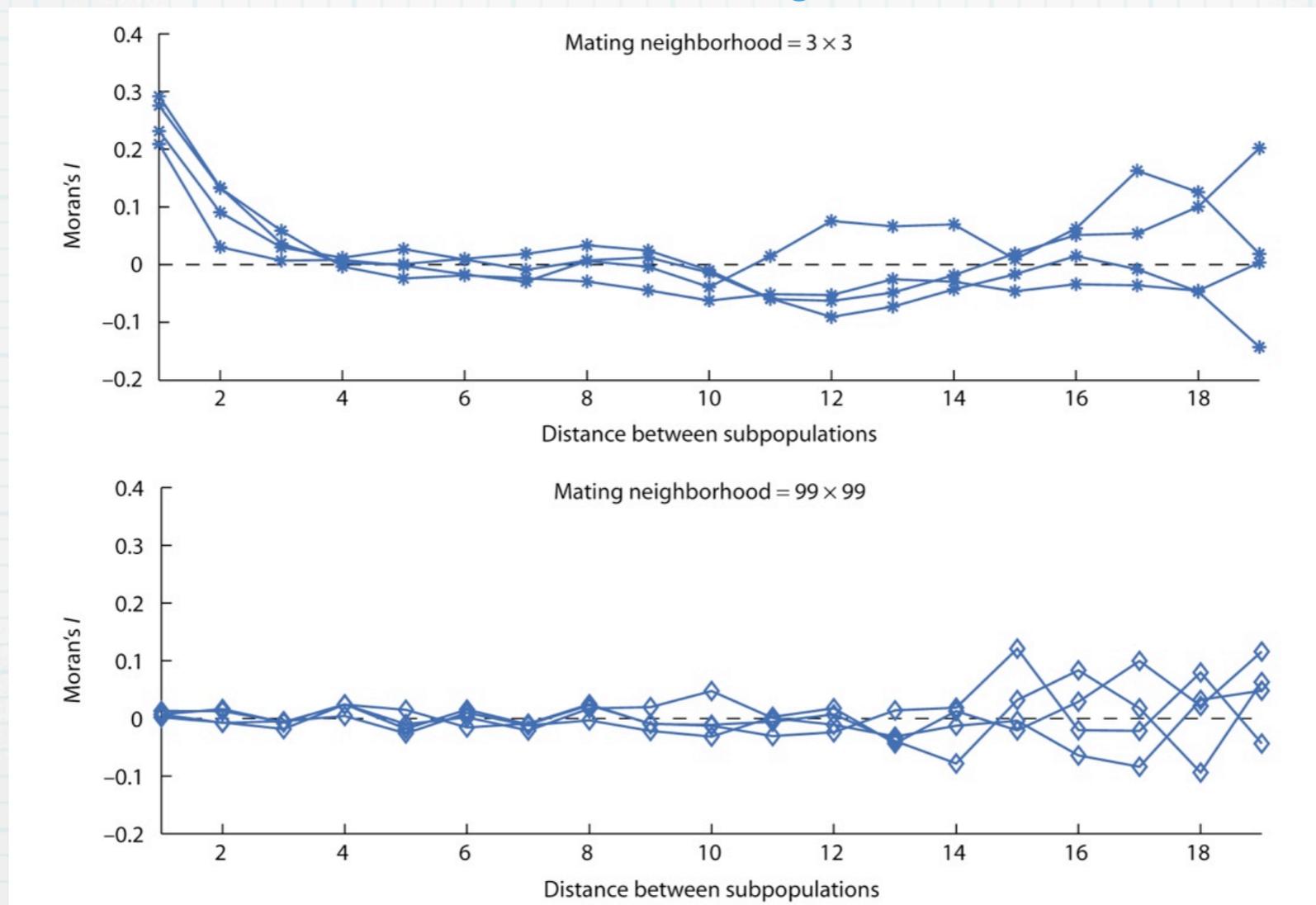
## Moran's I

- \* One classic statistic used to estimate spatial genetic structure is a correlation measure called **Moran's I**:

$$I_k = \frac{n \sum_{i=1}^n \sum_{j=1, i \neq j}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{W_k \sum_{i=1}^n (y_i - \bar{y})^2}$$

where  $k$  represents a distance class so that  $w_{ij}$  equals 1 if the distance between locations  $i$  and  $j$  equals  $k$ , and 0 otherwise. Within a distance class  $k$ ,  $n$  is the number of populations,  $y$  is the value of a genetic variable such as allele frequency for location  $i$  or  $j$ , and  $W_k$  is the sum of the weights  $w_{ij}$ , or  $2nk$ .

# (1) Genetic Populations



A  $100 \times 100$  grid was simulated for 200 generations and was then divided into square populations of  $10 \times 10$  individuals. The frequency of the A allele within each subpopulation is  $y_i$ . The distance classes are the number of subpopulations that separate pairs of subpopulations. Each line is based on an independent simulation of the  $100 \times 100$  population.

# (1) Genetic Populations

- \* Population structure has profound implications for genotype and allele frequencies
- \* Subdivision breaks up a population into smaller units that are each genetically independent to some degree
- \* Each deme may quickly reach fixation or loss, but both alleles can be maintained in the overall population since half of the subpopulations are expected to reach fixation and half loss for a given allele

# (1) Genetic Populations

- \* Processes that cause population structure can be thought of as both **creative** and **constraining** in evolutionary change
- \* **creative**: genetic isolation allows subpopulations to evolve independent allele frequencies and maintain unique alleles as is required for genetic adaptation to local environments under natural selection
- \* **constraining**: genetic isolation can prevent novel and even advantageous alleles from spreading throughout a population

# (1) Genetic Populations

## Migration vs. Gene Flow

- \* Migration is the movement of individuals from one place to another; it may or may not result in gene flow
- \* Gene flow requires that migrating individuals successfully contribute alleles to the mating pool of populations they join or visit
- \* Migration does not necessarily result in gene flow, and gene flow may occur without migration (e.g., plants' movement of pollen grains)

## (2) Direct Measures of Gene Flow

- \* **Parentage analyses**, using molecular genetic markers, identify the unknown parent(s) of a sample of progeny, and hence are considered direct measures of gene flow
- \* The **first step** in a parentage analysis is to examine the genotype of an individual progeny and its known parent for allelic matches
- \* At each locus, one of the alleles found in the known parent is observed in the progeny genotype

## (2) Direct Measures of Gene Flow

- \* The **second step** is to scan the genotypes of the candidate parents to see whether there is any individual with the other haplotype of the progeny
- \* All candidate parents that have a matching haplotype are possible fathers
- \* There is the possibility of **false positives**: an individual possesses a genotype with the same haplotype as the true parent

## (2) Direct Measures of Gene Flow

- \* Evaluating the chance of a false positive requires one to determine the chance of such a random match
- \* If  $p_i$  is the frequency of an allele at locus  $i$  in the matching haplotype, then the chance of a random match at locus  $i$  is  $P(\text{random match}) = p_i^2 + 2p_i(1-p_i)$
- \* Assuming independence among loci, then

$$P(\text{multilocus random match}) = \prod_{i=1}^{\text{loci}} (p_i^2 + 2p_i(1 - p_i))$$

## (2) Direct Measures of Gene Flow

- \* The ability to distinguish true parentage from apparent parentage due to random matches depends on both the allele frequencies at each locus as well as the total number of loci available
- \* Random matches become less likely as allele frequencies decrease and as the number of independent loci increases

## (2) Direct Measures of Gene Flow

- \* The probability that a haplotype matching the true parent will occur at random in a sample of  $n$  candidate parents is

$$P(\text{random match in } n \text{ individuals}) = 1 - (1 - P(\text{random match}))^n$$

- \* This probability can be thought of as the chance that a candidate parent is mistakenly assigned the true parent since its genotype provides the matching haplotype by chance, while the true parent is not identified since it is not included in the sample of candidate parents. This phenomenon is referred to as **cryptic gene flow** in paternity analysis

## (3) Fixation Indices

- \* The parentage analyses can be carried out when genotype data are available for a sample of candidate parents and a sample of progeny
- \* An alternative situation is where genotype data are determined for individuals sampled within and among a series of demographic locations
- \* For this situation, the fixation index can be extended

# (3) Fixation Indices

- \* In the case of multiple subpopulations, there can be deviations from HW expected frequencies of heterozygotes at two levels:
  - \* within each subpopulation due to non-random mating, and
  - \* among subpopulations due to population structure

# (3) Fixation Indices

The mathematical and biological definitions of heterozygosity for three levels of population organization. In the summations,  $i$  refers to each subpopulation  $1, 2, 3, \dots, n$  and  $p_i$  and  $q_i$  are the frequencies of the two alleles at a diallelic locus in subpopulation  $i$

$$H_I = \frac{1}{n} \sum_{i=1}^n \hat{H}_i$$

The average observed heterozygosity within each subpopulation

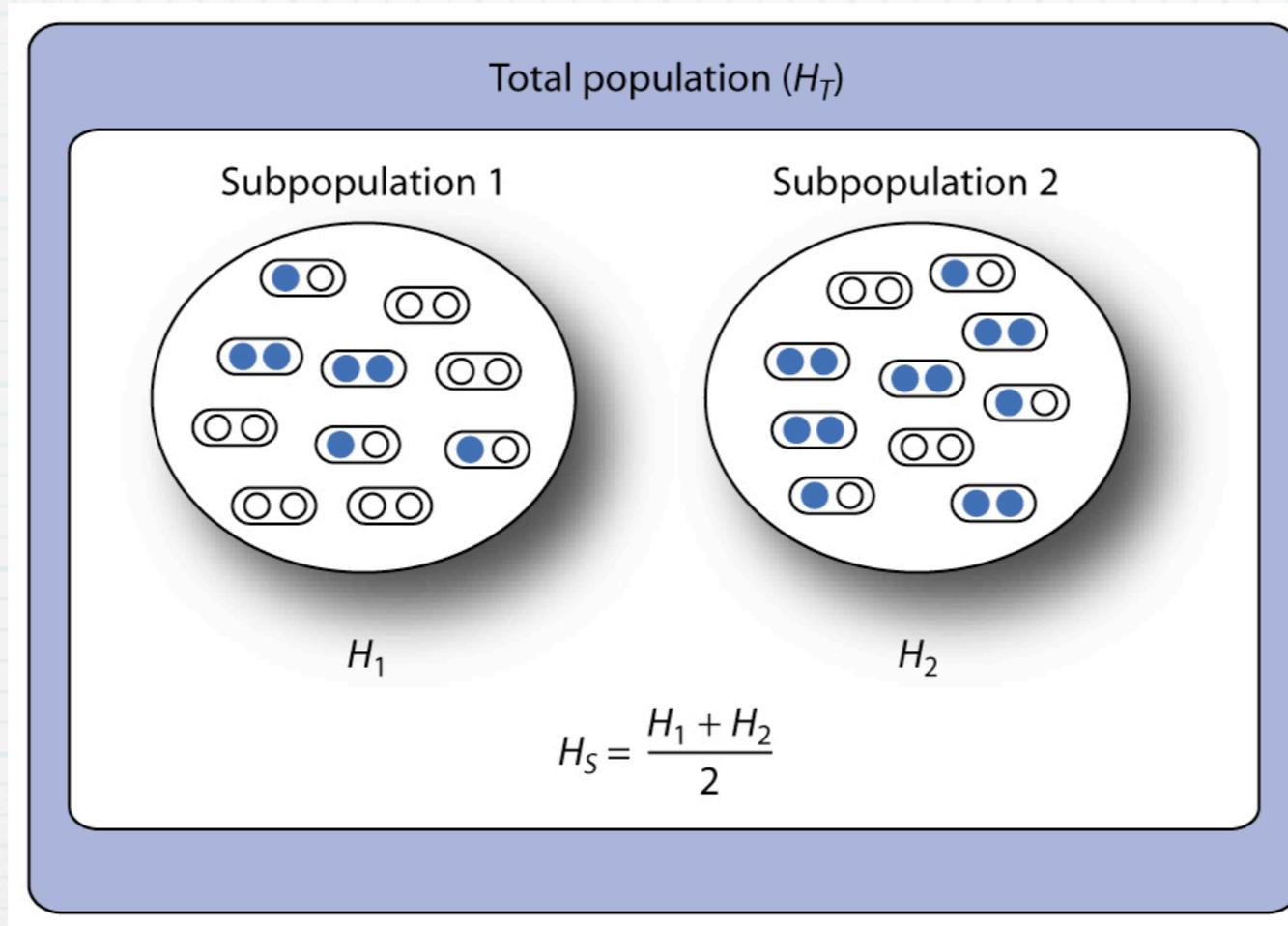
$$H_S = \frac{1}{n} \sum_{i=1}^n 2p_i q_i$$

The average expected heterozygosity of subpopulations assuming random mating within each subpopulation

$$H_T = 2\overline{pq}$$

The expected heterozygosity of the total population assuming random mating within subpopulations and no divergence of allele frequencies among subpopulations

# (3) Fixation Indices



$$\hat{H}_1 = 0.3 \quad \hat{H}_2 = 0.3 \quad H_I = \frac{1}{2}(0.3 + 0.3) = 0.3$$

$$p_1 = 13/20 \quad p_2 = 7/20 \quad H_S = \frac{1}{2}[2(0.65)(0.35) + 2(0.35)(0.65)] = 0.455$$

$$\bar{p} = (0.65 + 0.35)/2 = 0.5 \quad \bar{q} = (0.35 + 0.65)/2 = 0.5 \quad H_T = 2\bar{p}\bar{q} = 0.5$$

## (3) Fixation Indices

- \* With multiple subpopulations there is a possible excess or deficit of heterozygotes due to non-random mating within subpopulations and a possible deficit of heterozygotes among subpopulations compared to panmixia
- \* In the latter case the fixation index will show how much allele frequencies have diverged among subpopulations due to processes that cause population structure compared with the idea of uniform allele frequencies among subpopulations expected with panmixia

# (3) Fixation Indices

- \* Accounting for non-random mating and divergence of subpopulation allele frequency necessitates several new versions of the fixation index

# (3) Fixation Indices

The mathematical and biological definitions of fixation indices for two levels of population organization

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

The average difference between observed and HW expected heterozygosity within each subpopulation due to non-random mating

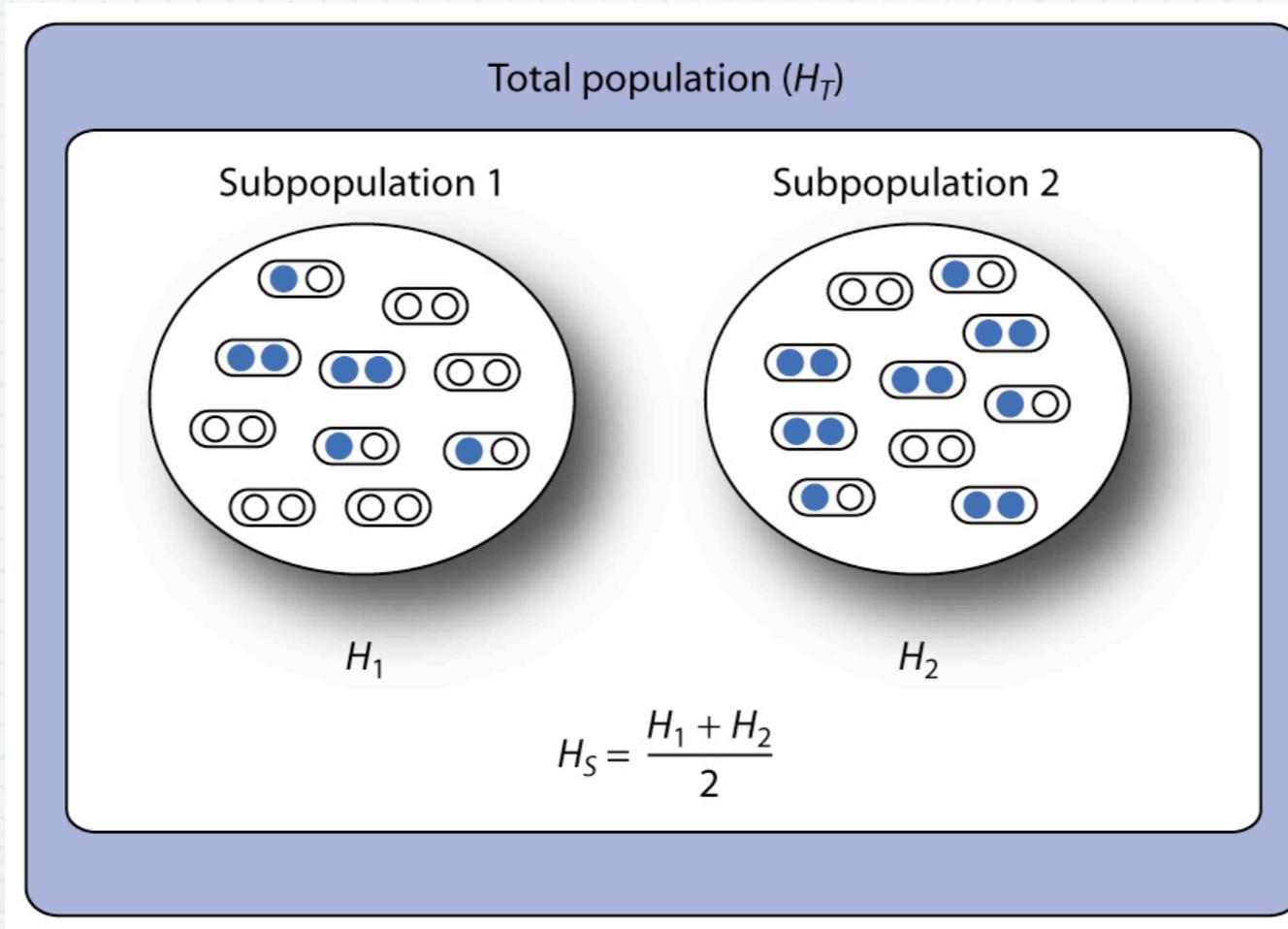
$$F_{ST} = \frac{H_T - H_S}{H_T}$$

The reduction in heterozygosity due to subpopulation divergence in allele frequency

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

The correlation between the states of two alleles in a genotype sampled at random from a single subpopulation given the possibility of non-random mating within populations and allele frequency divergence among populations

# (3) Fixation Indices



$$F_{IS} = \frac{H_S - H_I}{H_S} = \frac{0.455 - 0.3}{0.455} = 0.341$$

$$F_{ST} = \frac{H_T - H_S}{H_T} = \frac{0.5 - 0.455}{0.5} = 0.09$$

$$F_{IT} = \frac{H_T - H_I}{H_T} = \frac{0.5 - 0.3}{0.5} = 0.4$$

## (3) Fixation Indices

- \* The individual, subpopulation, and total population heterozygosities are identical in populations after compensating for the degree to which observed and expected heterozygosities are not met at different levels of population organization

## (3) Fixation Indices

- \* The average observed heterozygosity is greater or less than the average expected heterozygosity for subpopulations:

$$H_I = H_S(1 - F_{IS})$$

to the extent that there is non-random mating ( $F_{IS} \neq 0$ )

## (3) Fixation Indices

- \* The average expected heterozygosity for subpopulations is less than the expected heterozygosity of the total population under panmixia:

$$H_s = H_T(1 - F_{ST})$$

to the extent that subpopulations have diverged allele frequencies ( $F_{ST} > 0$ )

# (3) Fixation Indices

- \* The total deviation from expected heterozygosity within and among subpopulations is then:

$$H_I = H_T(1 - F_{IT})$$

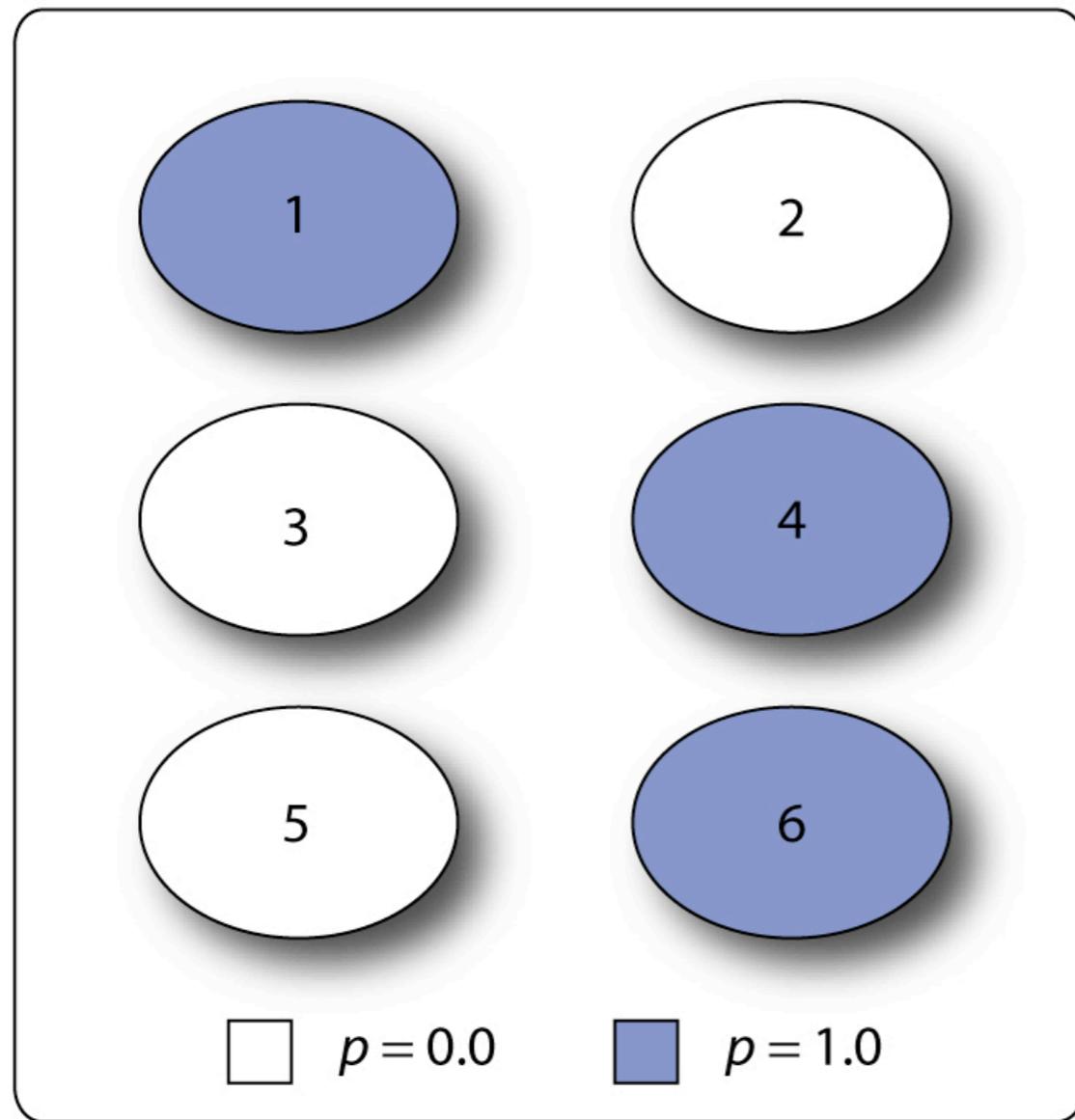
to the extent that subpopulations have diverged allele frequencies ( $F_{ST} > 0$ )

## (3) Fixation Indices

- \* Each fixation index expresses the degree to which random mating expectations for the frequency of heterozygous genotypes are not met
- \* Using the last three equations, it is possible to show how the total reduction in heterozygosity relates to the combined fixation due to non-random mating and subpopulation divergence:

$$1-F_{IT}=(1-F_{ST})(1-F_{IS})$$

# (3) Fixation Indices

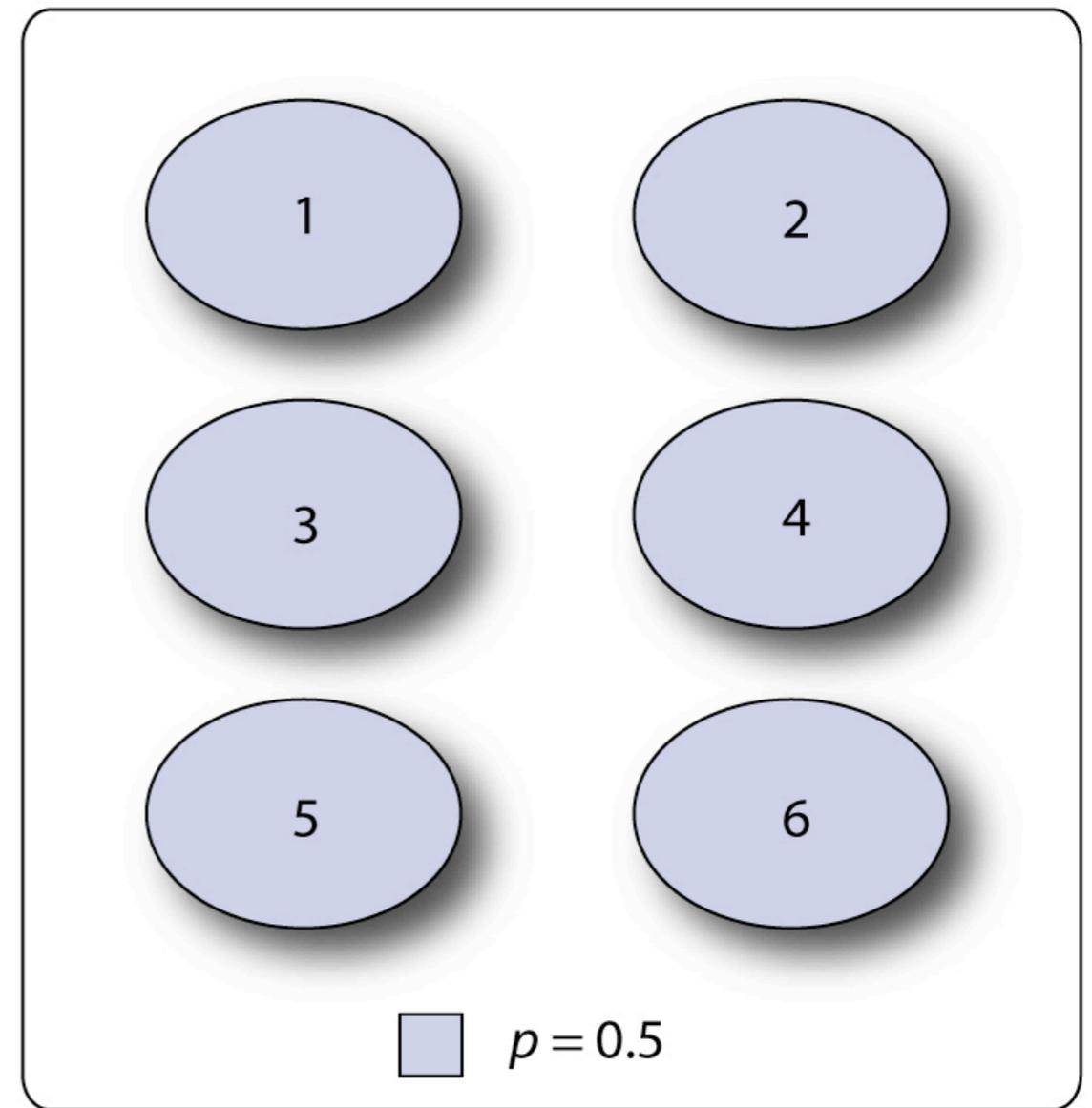


$$\bar{p} = 0.5$$

$$H_T = 2\bar{p}q = 0.5$$

$$H_S = 0$$

$$F_{ST} = 1$$



$$\bar{p} = 0.5$$

$$H_T = 2\bar{p}q = 0.5$$

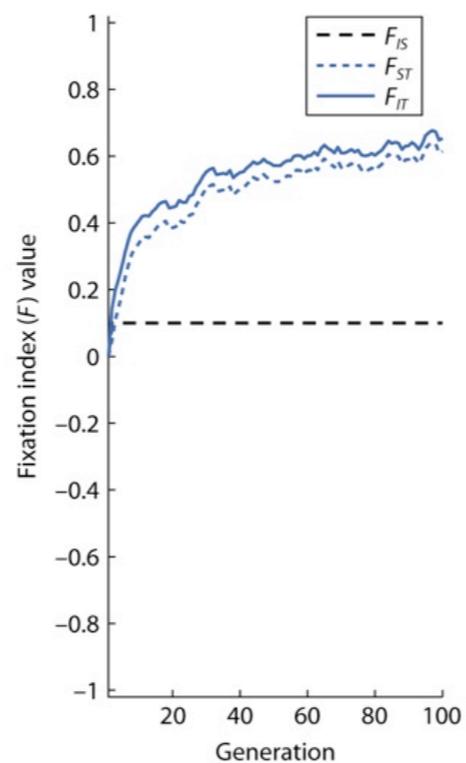
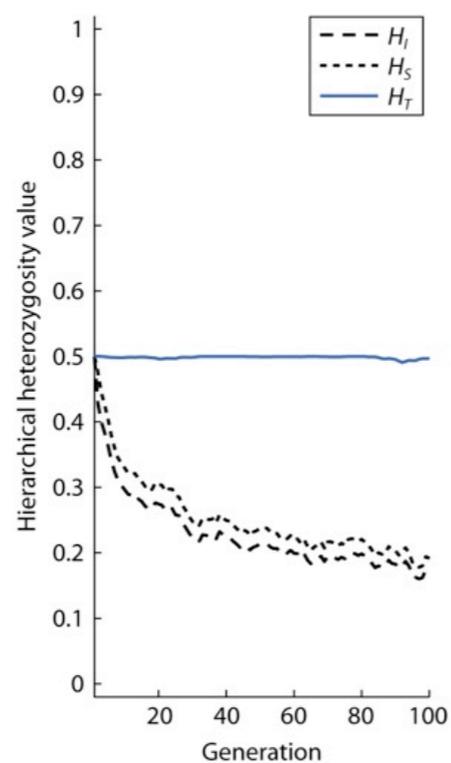
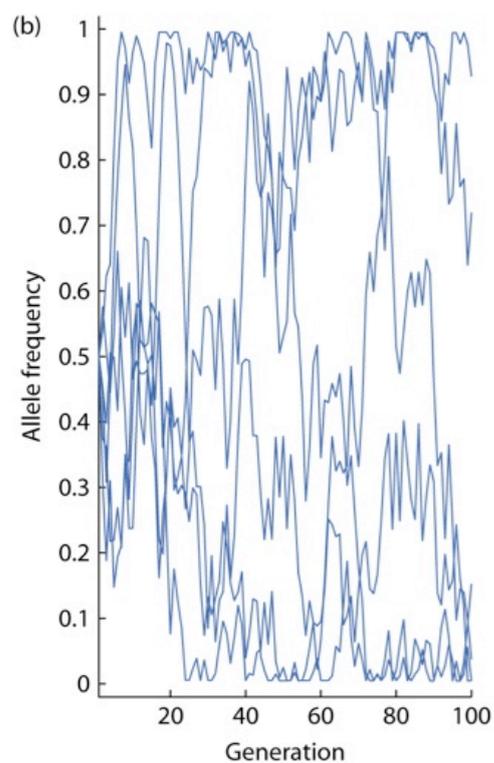
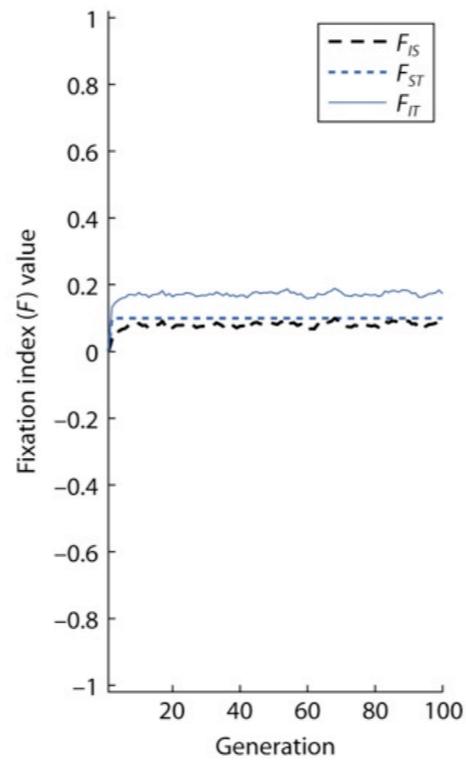
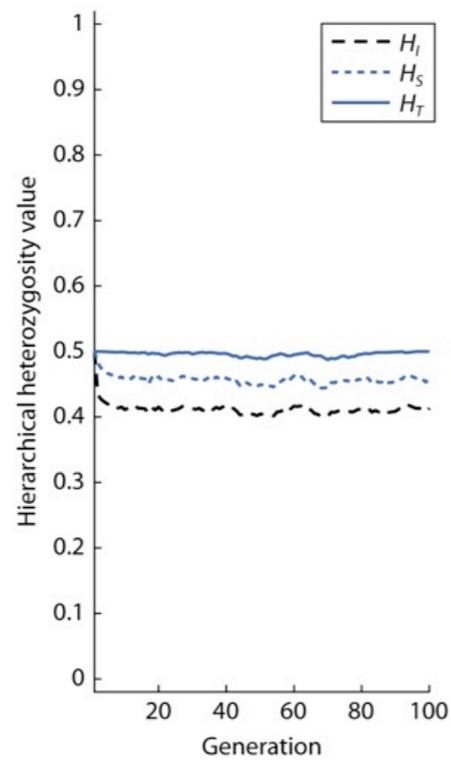
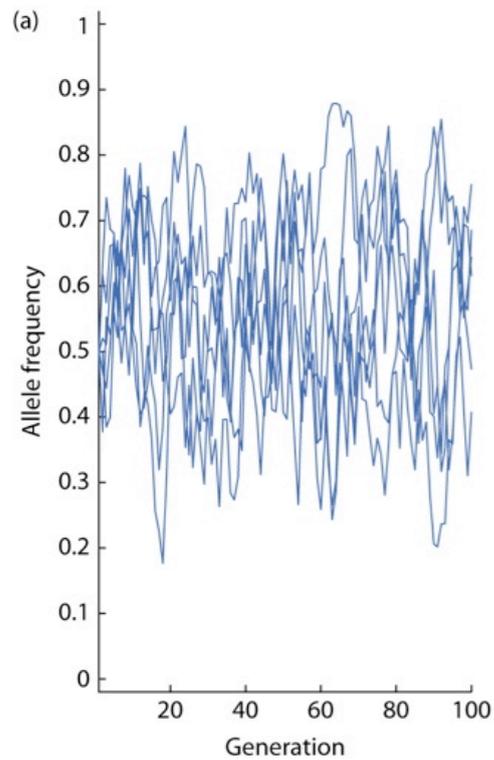
$$H_S = 0.5$$

$$F_{ST} = 0$$

## (3) Fixation Indices

- \* The relationship between allele frequencies of subpopulations, the hierarchical measures of heterozygosity, and the fixation indices can be seen in a simulation of gene flow and genetic drift in a subdivided population

# (3) Fixation Indices



- \* Each population contains 10 individuals
- \* The rate of gene flow is  $m=0.2$  in (a) and  $m=0.01$  in (b)
- \* The allele frequencies are shown for six randomly chosen subpopulations out of the 200 in the simulation
- \* The heterozygosities and fixation indices are calculated from all 200 subpopulations

# (4) Population Subdivision and the Wahlund Effect

- \* The **Wahlund effect** refers to the decreased expected frequency of heterozygotes in subpopulations with diverged allele frequencies compared with the expected frequency of heterozygotes in a panmictic population of the same total size with the same average allele frequencies

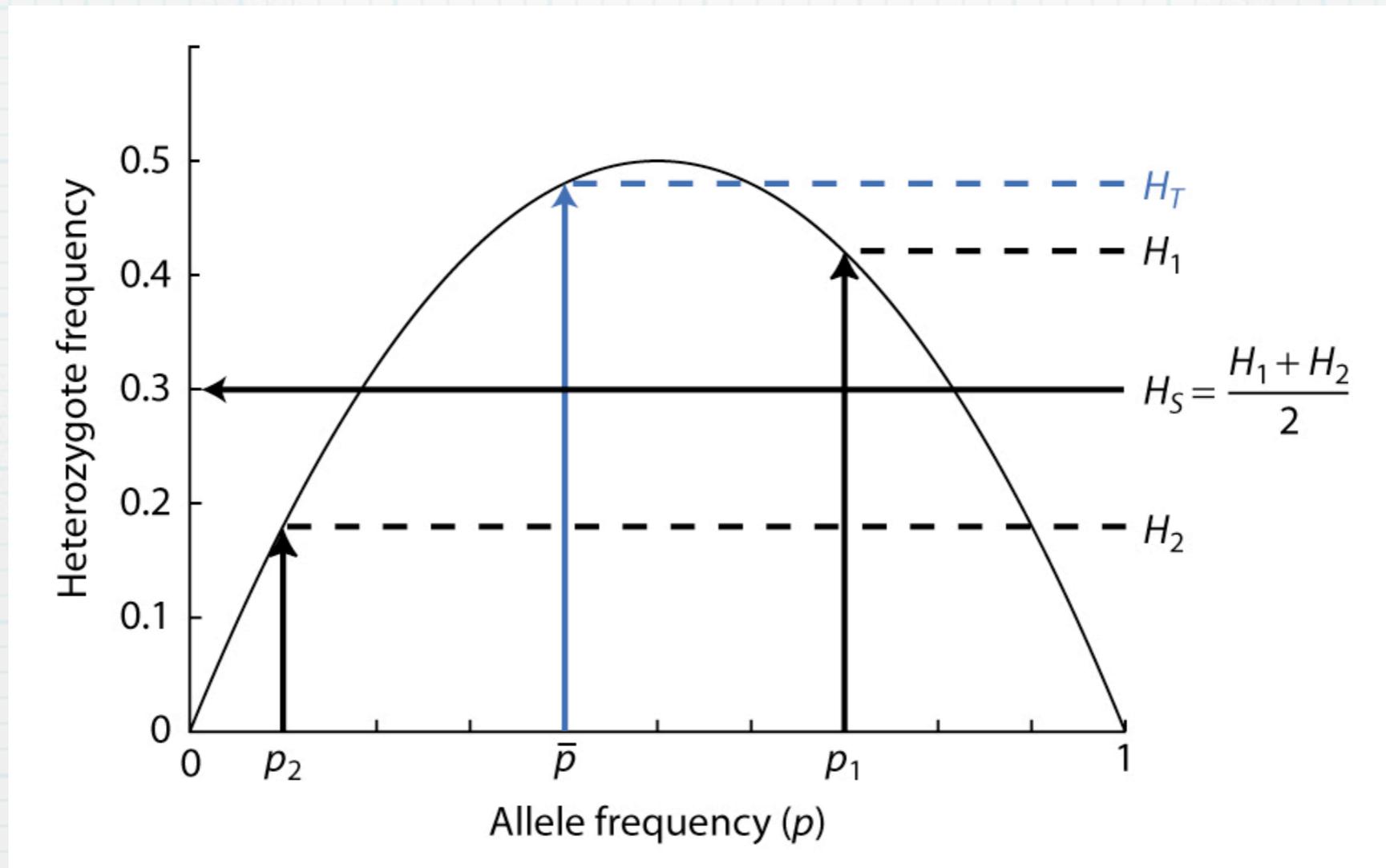
# (4) Population Subdivision and the Wahlund Effect

- \* Consider the case of a diallelic locus for two randomly mating demes
- \* The expected heterozygosity for each deme is  $H_i = 2p_iq_i$ , where  $i$  indicates a single subpopulation
- \* The average heterozygosity of the two demes is  $H_s = (2p_1q_1 + 2p_2q_2)/2$
- \* In contrast, the heterozygosity in the total population is  $H_T = 2\overline{pq}$

# (4) Population Subdivision and the Wahlund Effect

- \*  $H_T$  and  $H_S$  both cannot exceed 0.5, the maximum heterozygosity for a diallelic locus
- \* In addition,  $H_S$  is an average of  $H_1$  and  $H_2$ , so when subdivided populations have different allele frequencies,  $H_S$  will always be less than the expected heterozygosity of the total population
- \* These conditions assure that  $H_T \geq H_S$  when there is random mating within the subpopulations

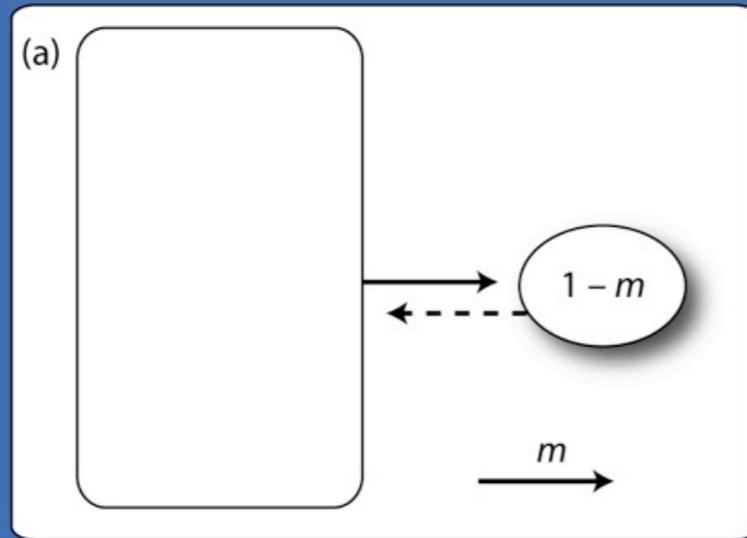
# (4) Population Subdivision and the Wahlund Effect



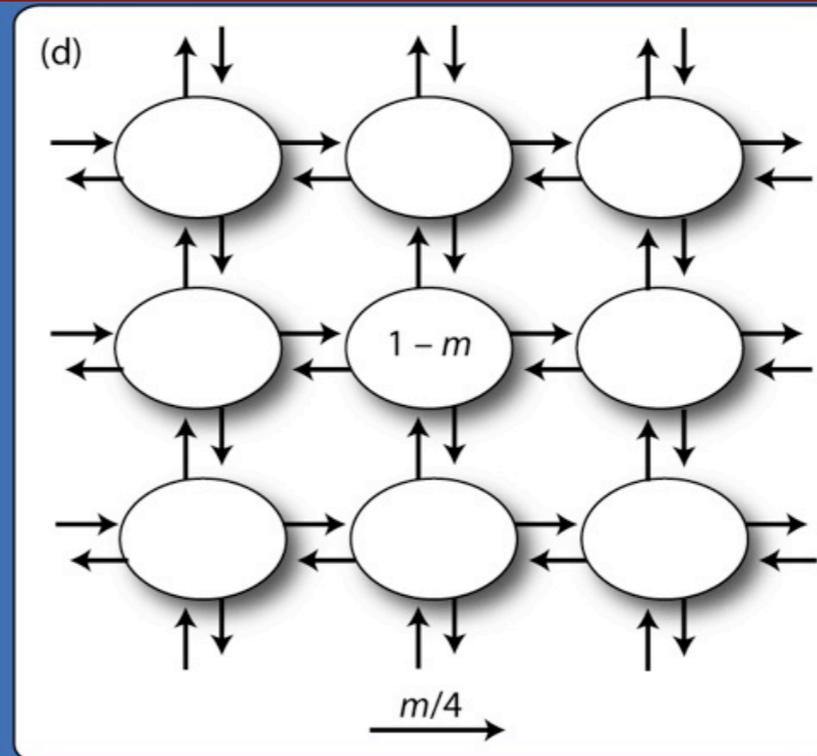
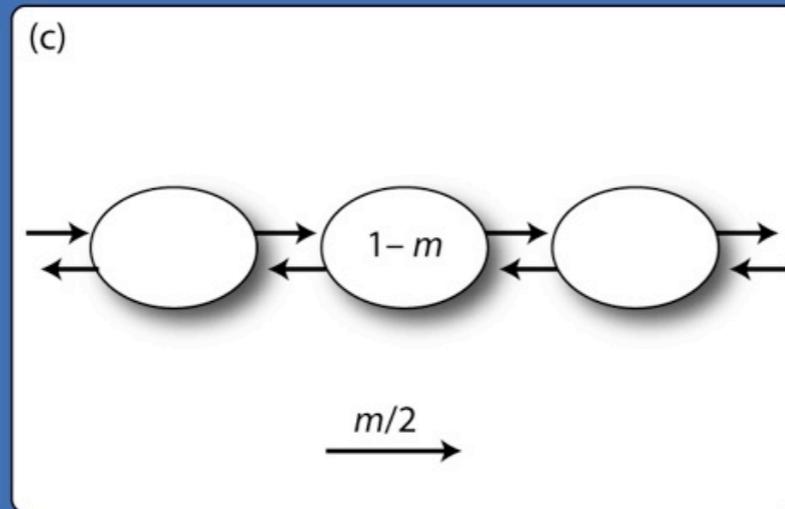
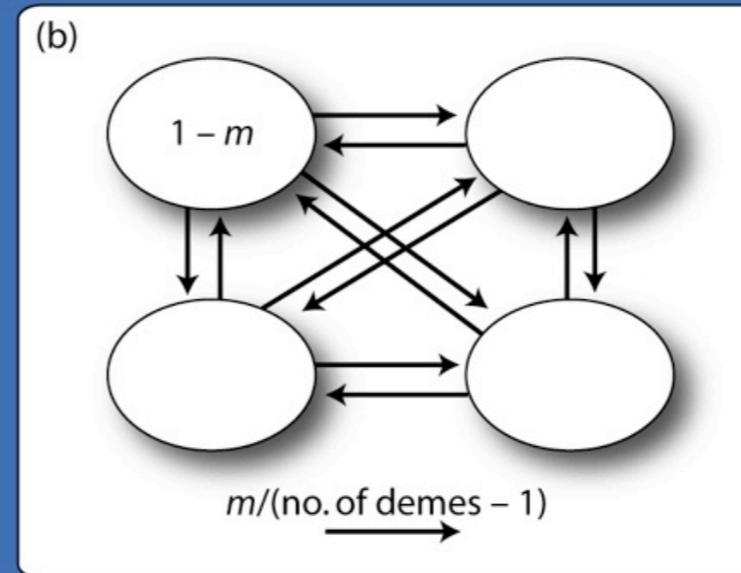
A graphical demonstration of the Wahlund effect for a diallelic locus in two demes

# (5) Models of Population Structure

The continent-island model



The island model



Stepping-stone models

# (5) Models of Population Structure

## The Continent-Island Model

- \* An idealized model of population subdivision and gene flow that assumes one very large population where allele-frequency changes only slowly over time connected by gene flow with a small population where migrants make up a finite proportion of the individuals present each generation. Gene flow from the island to the continent occurs but the continent population is assumed to be large enough that immigration has a negligible effect on allele frequencies.

# (5) Models of Population Structure

## The Continent-Island Model

- \* The island population experiences the replacement of a proportion  $m$  of its individuals through migration, with  $1-m$  of the original individuals remaining each generation

# (5) Models of Population Structure

## The Continent-Island Model

- \* Allele frequency (at a diallelic locus) in the island population one generation in the future is a function of (i) the allele frequency in the proportion of the island population that are not migrants, and (ii) the allele frequency in the proportion of the island population that arrives via gene flow from the continent population

$$p_{t=1}^{island} - p_{t=0}^{island} = -m(p_{t=0}^{island} - p^{continent})$$

# (5) Models of Population Structure

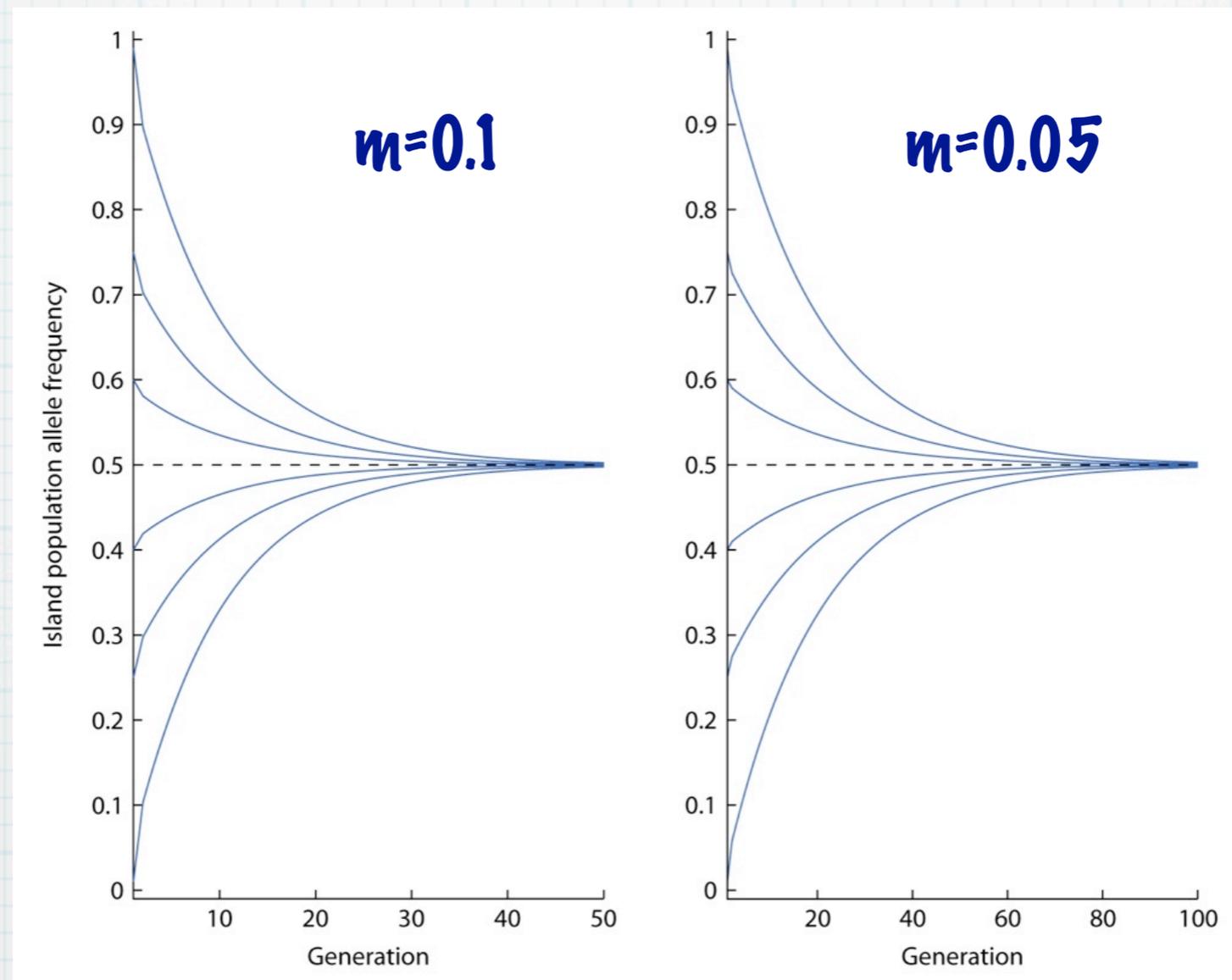
## The Continent-Island Model

- \* The equation can be generalized to  $t$  generations in the future:

$$p_t^{island} = p^{continent} + (p_{t=0}^{island} - p^{continent})(1 - m)^t$$

# (5) Models of Population Structure

## The Continent-Island Model



Allele frequency in the island population for a diallelic locus under the continent-island model. The continent population has an allele frequency of 0.5.

# (5) Models of Population Structure

## The Two-Island Model

- \* If we take the two populations in the continent-island model to be of equal size, removing the assumption that one population serves as an unchanging source of migrants, we obtain the two-island model
- \* Two rates of gene flow:  $m_1$  and  $m_2$
- \* Assuming  $m_1 = m_2$ , the allele frequency in either subpopulation is

$$p_t = \bar{p} + (p_{t=0} - \bar{p})(1 - m)^t \quad (\bar{p} = (p_1 + p_2)/2)$$

# (5) Models of Population Structure

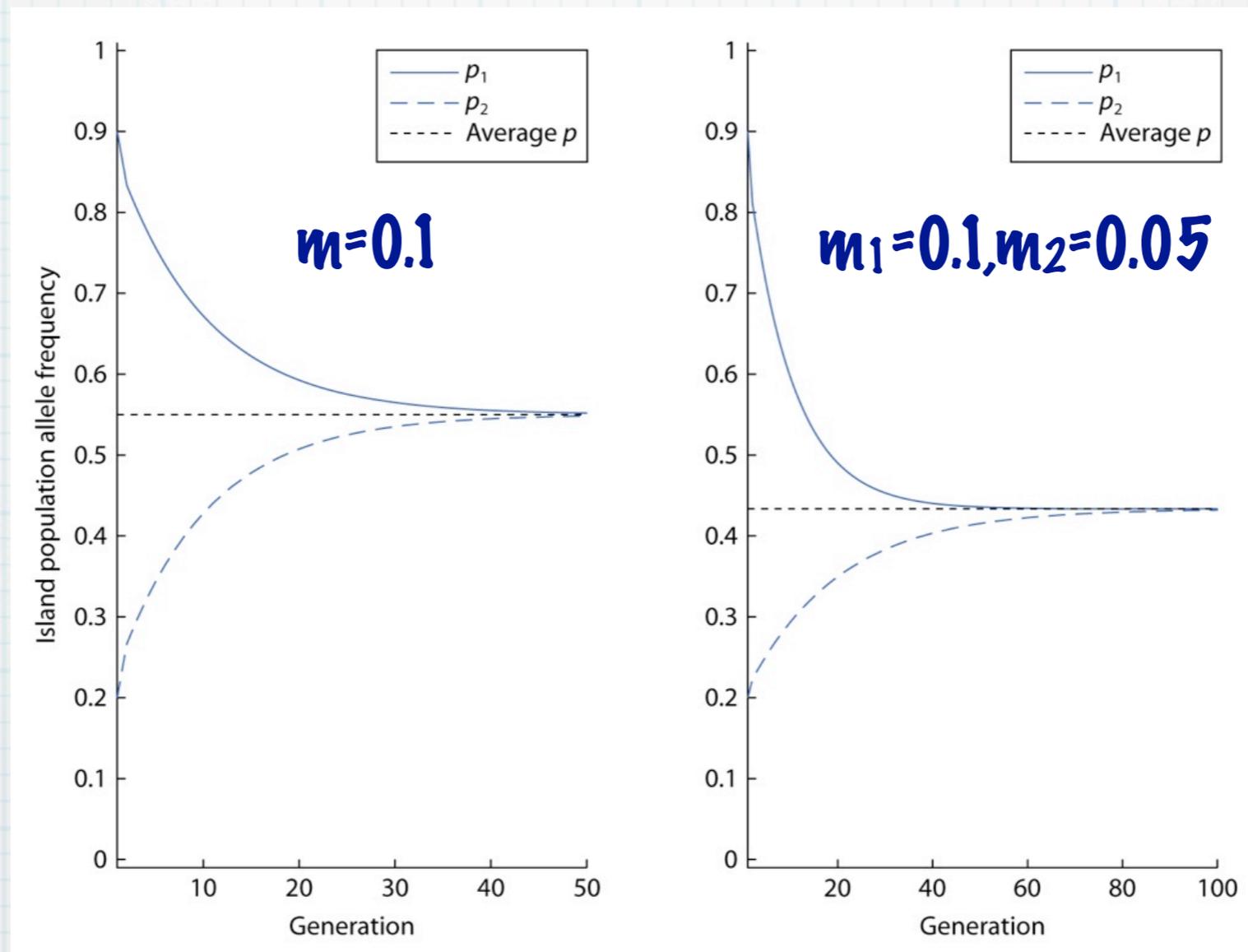
## The Two-Island Model

- \* When the rates of gene flow are not equal, then the average allele frequency is

$$\bar{p} = \frac{p_1 m_2 + p_2 m_1}{m_1 + m_2}$$

# (5) Models of Population Structure

## The Two-Island Model



Allele frequency in the two-island model for a diallelic locus.

# (5) Models of Population Structure

## The Infinite-Island Model

- \* An idealized model of population subdivision and gene flow that assumes an infinite number of identical subpopulations and that each subpopulation experiences an equal probability of gene flow from all other subpopulations
- \* It is most commonly assumed that the total population is made up of an infinite set of subpopulations each of size  $N_e$  with  $m$  percent of each subpopulation's gene copies exchanged at random with the rest of the population every generation

# (5) Models of Population Structure

## The Infinite-Island Model

- \* The expected value of the fixation index for subpopulations compared to the total population is:

$$F_{ST} = 1 - e^{-\frac{1}{2N_e}t}$$

# (5) Models of Population Structure

## The Infinite-Island Model

- \* Accounting for both gene flow and genetic drift, the fixation index is:

$$F_t = \frac{1}{2N_e}(1 - m)^2 + \left(1 - \frac{1}{2N_e}\right) F_{t-1}(1 - m)^2$$

- \* This equation is an expression for the balance of gene flow and genetic drift among multiple subpopulations so  $F$  is identical to  $F_{ST}$

# (5) Models of Population Structure

## The Infinite-Island Model

- \* With the assumption that the migration rate is small and much, much less than the effective population size, an approximation for the expected amount of fixation among subpopulations at equilibrium in an infinite island population is:

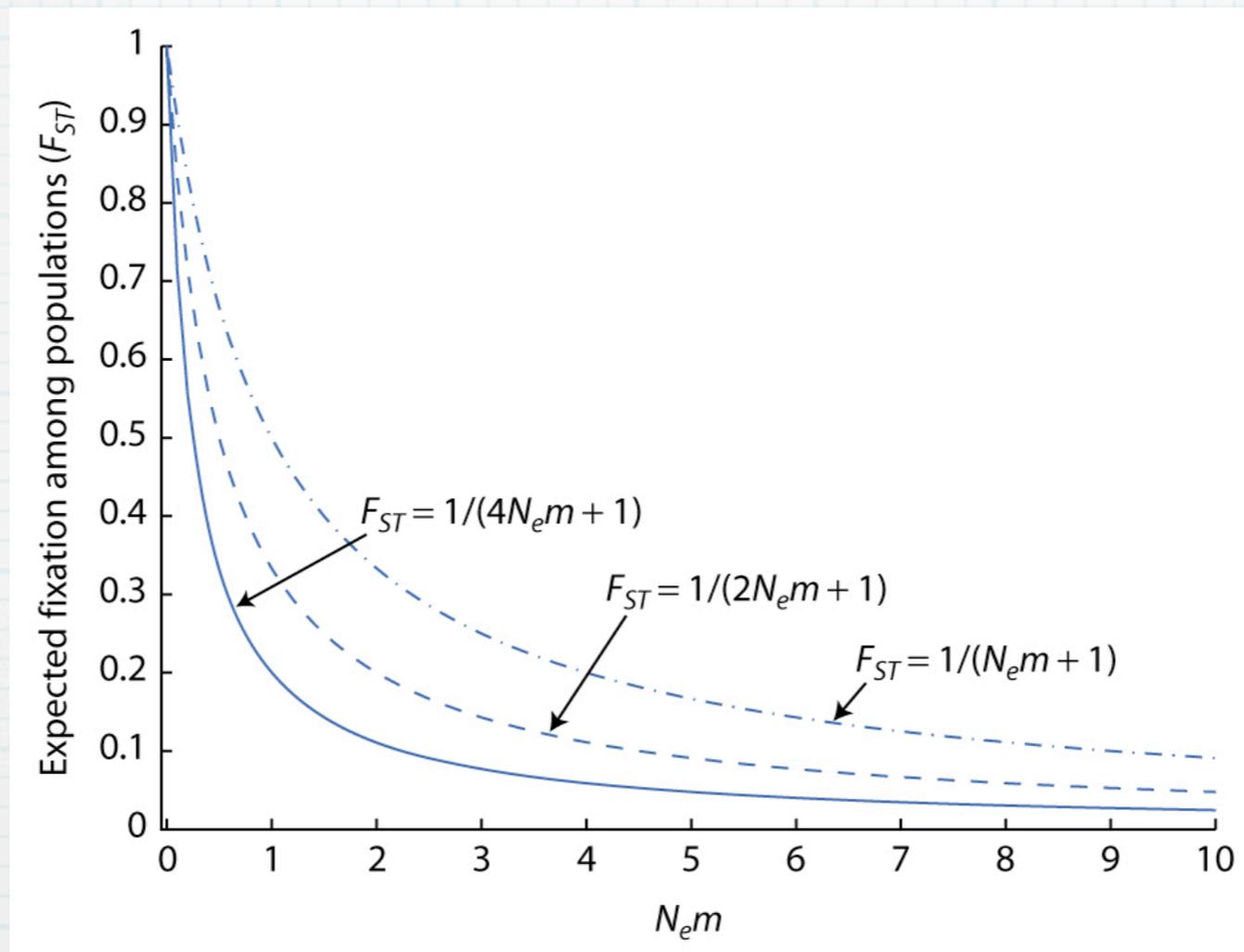
$$F_{ST} \approx \frac{1}{4N_e m + 1}$$

- \* To get an expected effective migration rate given an amount of genetic differentiation among subpopulations in the infinite island model:

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

# (5) Models of Population Structure

## The Infinite-Island Model



Expected levels of fixation among subpopulations depend on the effective migration rate ( $N_e m$ ). Each curve represents expected  $F_{ST}$  for loci with different probabilities of autozygosity (from bottom to top:  $1/(2N_e)$ ,  $1/N_e$ , and  $2/N_e$ ).

# (5) Models of Population Structure

## The Infinite-Island Model

- \* The expected relationship between the fixation index and the effective number of migrants relies on the infinite island model for two reasons:
  1. all subpopulations have an identical rate of migration from all other subpopulations, so there is only a single migration rate ( $m$ ) that applies to all subpopulations
  2. Since there are an infinite number of subpopulations, the entire ensemble population will never reach fixation or loss due to genetic drift

# (5) Models of Population Structure

## The Infinite-Island Model

- \* In a **finite island model**, the entire set of subpopulations will eventually reach fixation or loss and  $F_{ST}$  will eventually decline to zero since the entire set of subpopulations will eventually reach fixation or loss due to genetic drift
- \* The expected amount of genetic differentiation in this model is, where  $n$  is the number of subpopulations is:

$$G_{ST} = \frac{1}{4N_e m \left( \frac{n}{n-1} \right)^2 + 1}$$

# (5) Models of Population Structure

## Stepping-stone and Metapopulation Models

- \* The **stepping-stone model** approximates the phenomenon of **isolation by distance** among discrete subpopulations by allowing most or all gene flow to be only between neighboring subpopulations
- \* Under this model, the correlation between allelic states decreases with increasing distance between subpopulations

# (5) Models of Population Structure

## Stepping-stone and Metapopulation Models

- \* An extension of the stepping-stone model is the **metapopulation model**, which approximates the continual extinction and recolonization seen in many natural populations in addition to the process of gene flow
- \* Gene flow in this model is of two types: (i) gene flow among existing subpopulations, and (ii) gene flow that occurs when an open patch is colonized to replace a subpopulation that went extinct

# (5) Models of Population Structure

## Stepping-stone and Metapopulation Models

- \* The pattern of gene flow that takes place during colonization may take different forms:
  - \* **migrant-pool gene flow**: colonists are sampled at random from all subpopulations
  - \* **propagule-pool gene flow**: colonists are sampled at random from only a single random subpopulation

# (5) Models of Population Structure

## Stepping-stone and Metapopulation Models

- \* The impact of the form of colonization on heterozygosity in newly established subpopulations within a metapopulation is described by

$$F_{ST}^{colony} = \frac{1}{2k} + \phi \left( 1 - \frac{1}{2k} \right) F_{ST}$$

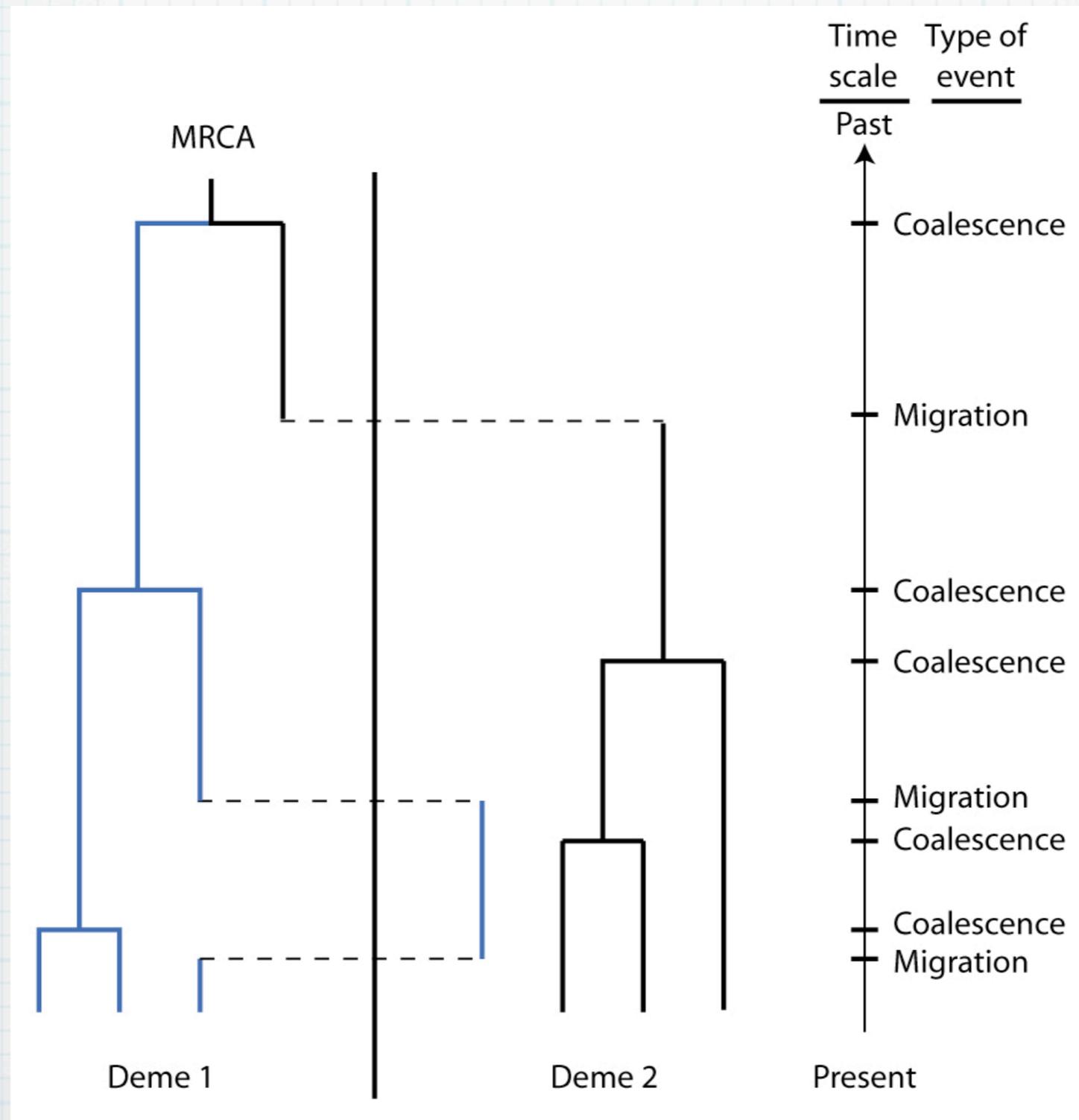
where  $F_{ST}^{colony}$  is the expected allele frequency differentiation in newly established subpopulations,  $k$  is the number of diploid colonists,  $F_{ST}$  is the degree of allele frequency differentiation among the existing subpopulations, and  $\phi$  is the probability that the two alleles in a newly established population come from the same or different subpopulations.

# (5) Models of Population Structure

## Stepping-stone and Metapopulation Models

- \* Colonization corresponds to the propagule pool for  $\Phi=1$  (the chance of one that two founding alleles come from the same subpopulation) and the migrant pool for  $\Phi=0$  (no chance that two founding alleles come from the same subpopulation, so all alleles must come from different subpopulations)

# (6) Population Structure and the Coalescent



# (6) Population Structure and the Coalescent

- \* To combine the coalescent with migration events, we will assume that both coalescent and migration events are rare so that when an event does occur going back in time it is either coalescence or migration

# (6) Population Structure and the Coalescent

- \* In a subdivided population, each generation there is the chance that a lineage in one deme migrates to some other deme
- \* The rate of migration,  $m$ , is the chance that a lineage migrates each generation
- \* The chance that  $t$  generations pass before a migration event occurs is

$$P(T_{\text{migration}}=t)=(1-m)^{t-1}m$$

# (6) Population Structure and the Coalescent

- \* The chance that a single lineage in any of the  $d$  demes migrates at generation  $t$  can be approximated as

$$P(T_{migration} = t) = e^{-t \frac{Md}{2}}$$

where  $M=4N_e m$

[each time unit is equivalent to  $2N_e d$  generations]

# (6) Population Structure and the Coalescent

- \* The total chance of migration for  $k$  ancestral lineages of the  $d$  demes is approximated as

$$P(T_{migration} = t) = e^{-t \frac{M d}{2} k}$$

# (6) Population Structure and the Coalescent

- \* The chance that one of  $k$  lineages migrates at or before a certain time can then be approximated with the cumulative exponential distribution

$$P(T_{migration} \leq t) = 1 - e^{-t \frac{M_d}{2} k}$$

# (6) Population Structure and the Coalescent

- \* When two independent processes (coalescence and migration) are operating, the genealogical model becomes one of following lineages back in time and waiting for an event to happen
- \* The total chance of any event is the sum of the independent probabilities for each type of mutually exclusive events

# (6) Population Structure and the Coalescent

- \* Since lineages cannot coalesce unless they are in the same deme, the chance of a coalescent event, when there are  $k_i$  ancestral lineages in deme  $i$ , is

$$P(T_{\text{coalescent}} \leq t) = 1 - e^{-td \sum_{i=1}^d \frac{k_i(k_i-1)}{2}}$$

# (6) Population Structure and the Coalescent

- \* The total chance of an event occurring when going back in time is then

$$P(T_{event} \leq t) = 1 - e^{-td \left[ k \frac{M}{2} + \sum_{i=1}^d \frac{k_i(k_i - 1)}{2} \right]}$$

# (6) Population Structure and the Coalescent

- \* When an event does occur, it is necessary to determine whether the event is a coalescence or a migration
- \* The total chance that the event is either a migration or a coalescence event is

$$d \left[ k \frac{M}{2} + \sum_{i=1}^d \frac{k_i(k_i - 1)}{2} \right]$$

# (6) Population Structure and the Coalescent

\* Therefore, the chance the event is a migration is

$$\frac{dk \frac{M}{2}}{d \left[ k \frac{M}{2} + \sum_{i=1}^d \frac{k_i(k_i-1)}{2} \right]} = \frac{kM}{k(M-1) + \sum_{i=1}^d k_i^2}$$

\* And, the chance that event is a coalescence is

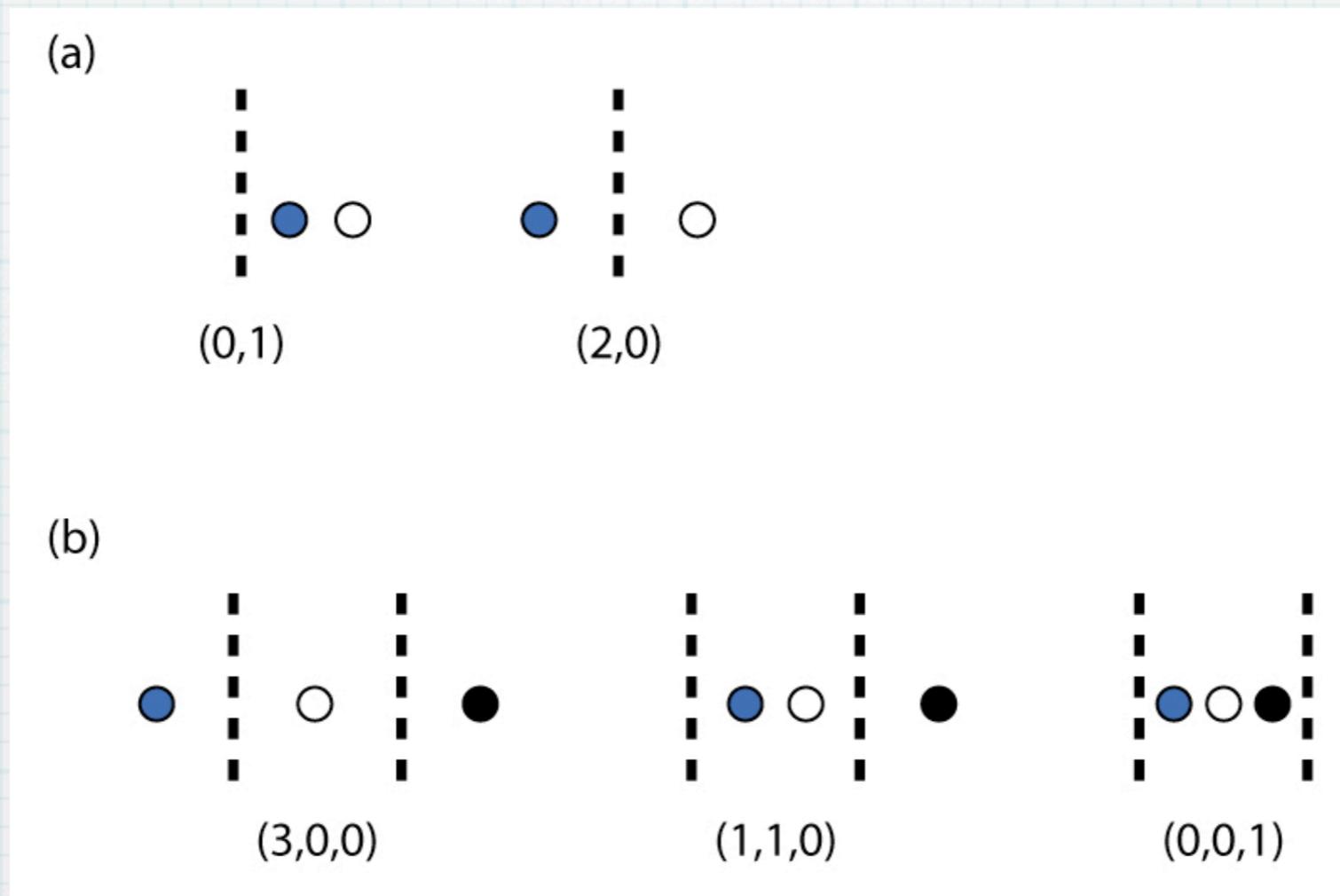
$$\frac{d \sum_{i=1}^d \frac{k_i(k_i-1)}{2}}{d \left[ k \frac{M}{2} + \sum_{i=1}^d \frac{k_i(k_i-1)}{2} \right]} = \frac{\sum_{i=1}^d (k_i^2 - k_i)}{k(M-1) + \sum_{i=1}^d k_i^2}$$

# (6) Population Structure and the Coalescent

- \* In the simple finite island model, the mean and variance of coalescent times for a sample of alleles can be calculated
- \* Obviously, both are dependent on how the  $n$  genes are sampled, that is, whether they are from the same deme or from different demes

# (6) Population Structure and the Coalescent

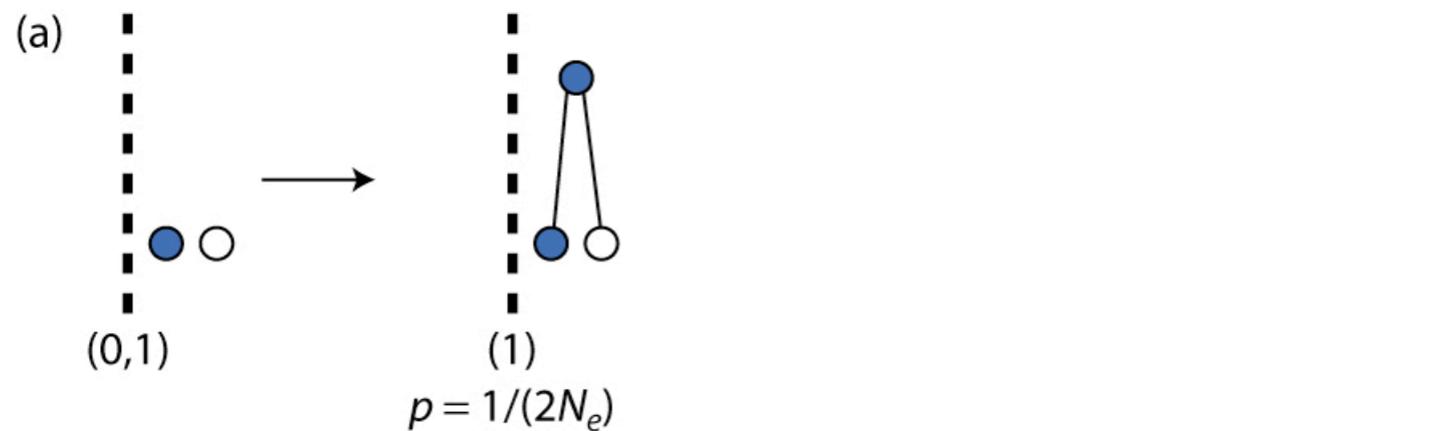
- \* Let the sample configuration be given as  $d=(d_1,d_2,\dots,d_n)$ , where  $d_i$  is the number of demes with  $i$  members



# (6) Population Structure and the Coalescent

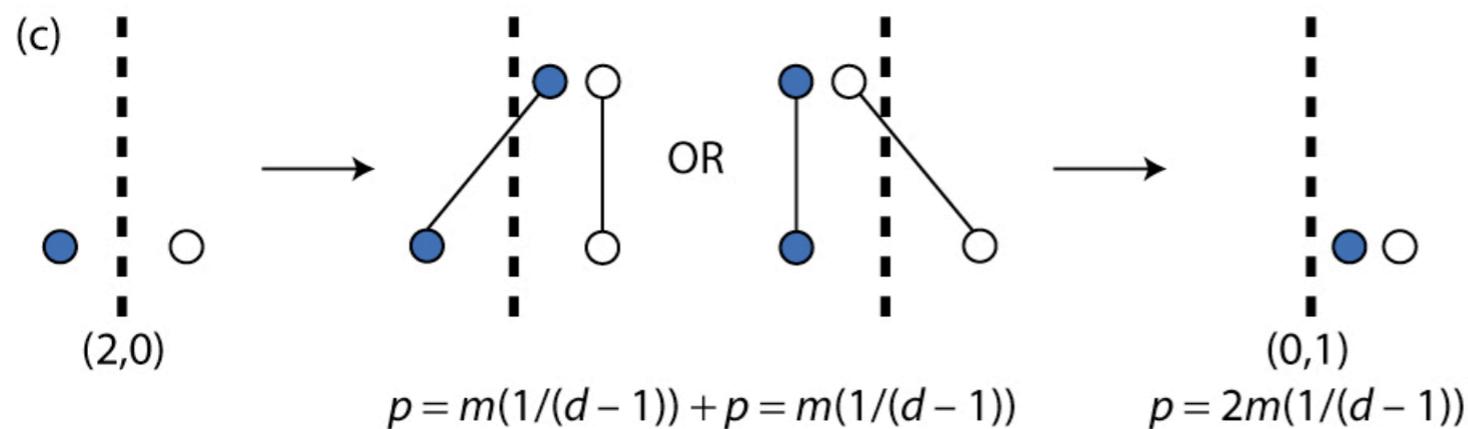
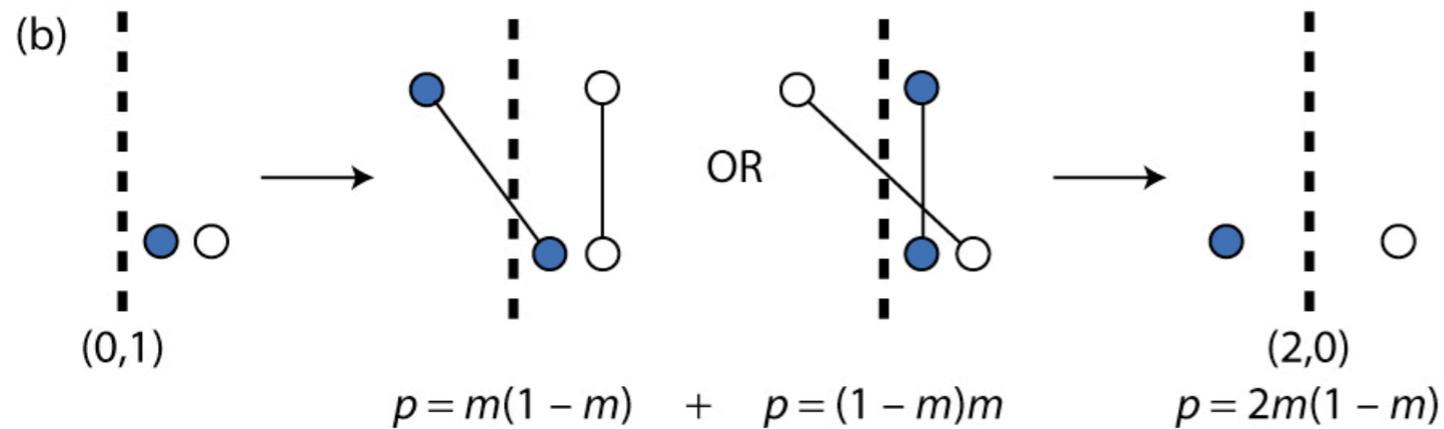
- \* In the beginning, we have  $n=(1d_1+2d_2+\dots+nd_n)$
- \* Subsequent events change  $d$  until the state  $(1,0,\dots,0)$  is reached
- \* Let  $T_d$  denote the time until the MRCA, given the sample configuration is  $d$
- \* In particular,  $T_{(2,0)}$  is the time for two alleles in two different demes to find an MRCA, and  $T_{(0,1)}$  is the time for two alleles in the same deme to find an MRCA

# (6) Population Structure and the Coalescent



The total chance that any event occurs in the previous generation is

$$2m + \frac{1}{2N_e}$$



The total chance of moving from configuration (2,0) to (0,1) is

$$2m \frac{1}{d-1}$$

# (6) Population Structure and the Coalescent

## The Average Length of a Genealogy With Migration

$$\bar{T}_{(0,1)} = \left( \frac{1}{2m + \frac{1}{2N_e}} \right) \left( \frac{1}{2N_e} \right) + \left( \frac{1}{2m + \frac{1}{2N_e}} \right) (2m) \bar{T}_{(2,0)}$$

The event is a coalescence

The event is a migration

The average time to coalescence for 2 lineages in the same deme

The average time of either a coalescence or a migration

The average time to coalescence for 2 lineages in different demes

$$\bar{T}_{(2,0)} = \frac{d-1}{2m} + \bar{T}_{(0,1)}$$

The average time of 2 lineages to migrate into the same deme

# (6) Population Structure and the Coalescent

The Average Length of a Genealogy With Migration

Solving the two equations gives

$$\bar{T}_{(0,1)} = 2N_e d$$

$$\bar{T}_{(2,0)} = 2N_e d + \frac{d-1}{2m}$$

# (6) Population Structure and the Coalescent

## The Average Length of a Genealogy With Migration

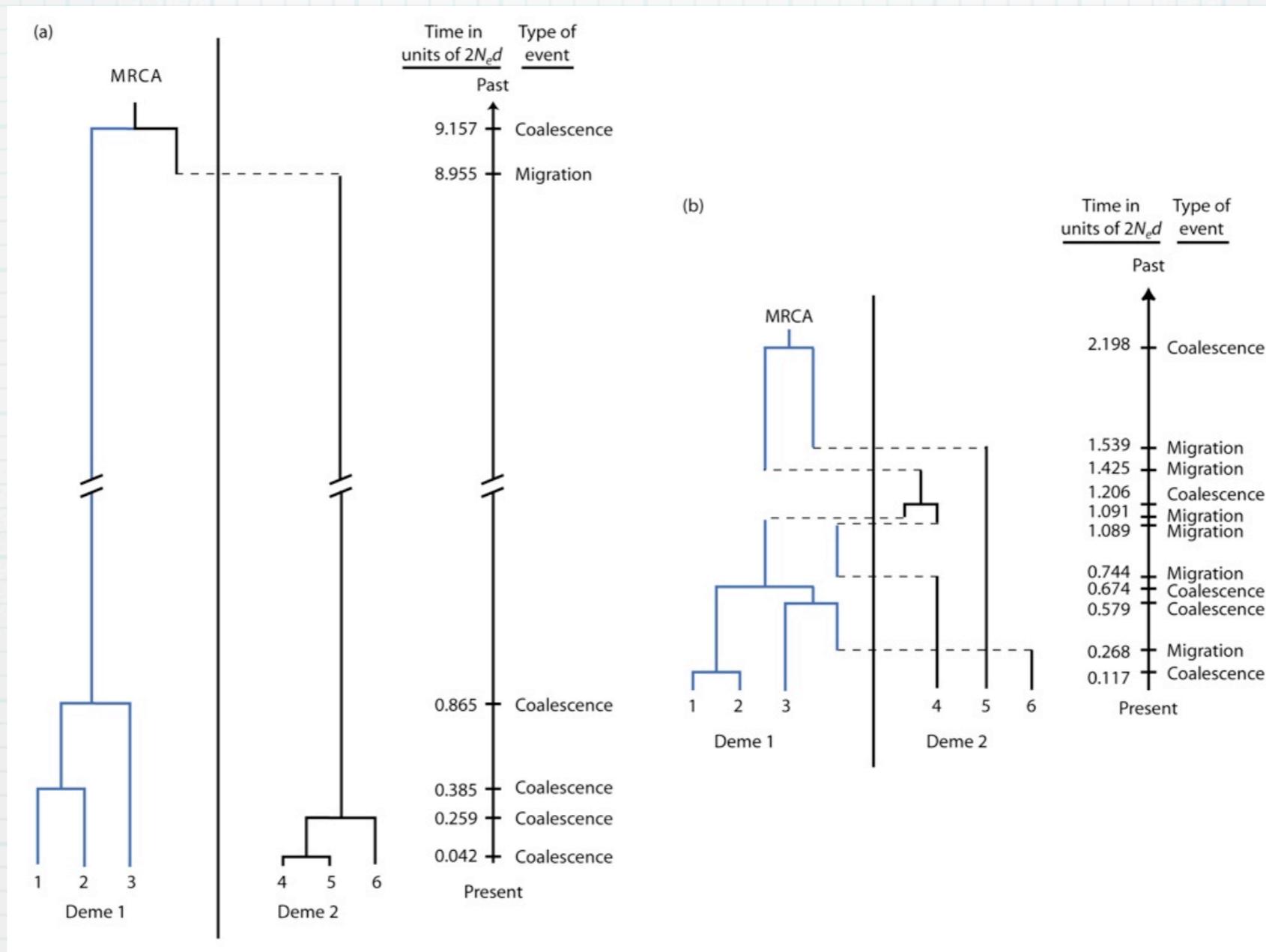
The average coalescence time for two lineages sampled at random from a subdivided population is

$$\bar{T} = \frac{1}{d} \bar{T}_{(0,1)} + \frac{d-1}{d} \bar{T}_{(2,0)} = 2N_e d + \frac{(d-1)^2}{2md}$$

Chance the 2 lineages are from the same deme

Chance the 2 lineages are from different demes

# (6) Population Structure and the Coalescent



(a)  $M=4N_e m=0.2$

(a)  $M=4N_e m=2.0$

Genealogies for 6 lineages initially divided evenly between two demes.

# (6) Population Structure and the Coalescent

## The Average Length of a Genealogy With Migration

Thinking about population structure as the difference in average coalescence times for a pair of lineages in the total population and a pair of lineages in a subpopulation, we obtain

$$F_{ST} = \frac{\bar{T} - \bar{T}_{(0,1)}}{\bar{T}}$$

# Summary

- \* Spatial and temporal separation of discrete subpopulations as well as isolation by distance in continuous populations both result in mating that is not random throughout a population
- \* Without enough gene flow to maintain random mating (panmixia), genetic drift causes a divergence of allele frequencies among subpopulations
- \* Parentage analyses use genotypes of progeny and one known parent to infer the haplotype of the unknown parent
- \* The Wahlund effect demonstrates that genetic variation can be stored as variance in allele frequencies among subpopulations or as heterozygosity within a panmictic population

# Summary

- \*  $F_{IS}$  measures the average excess or deficit of heterozygous genotypes compared with random mating
- \*  $F_{ST}$  measures the deficit of heterozygosity in subpopulations due to population structure compared to heterozygosity expected with panmixia
- \*  $F_{IT}$  measures the total excess or deficit of heterozygous genotypes due to both non-random mating within and allele-frequency divergence among subpopulations
- \* The coalescent model can be combined with migration events to predict average times to coalescence in divided populations