

EEB 70901 Population Genetics

Lecture 4:
Population Subdivision

PCA analyses, population assignments, and admixture analyses are very powerful methods for clustering individuals and exploring the genetic relationships among individuals for large SNP data sets. But the results are not easily interpretable in terms of population genetics. These methods are agnostic regarding the processes causing differences among individuals. They do not provide estimates of divergence times or migration rates, and cannot be used directly to infer the demographic history of the populations analyzed. They find their strength when populations are not well defined and exploratory analyses are needed to define natural units for population genetic analyses.



about PCA, assignment, admixture
and delimitation methods in general

Monty Slatkin aka “the oracle”

Big Picture - review

So far we have covered:

How allele and genotype frequencies change:

- under the simplest of models (Hardy-Weinberg)
- allowing for genetic drift

How gene genealogies describe the ancestry of a sample and can predict sequence differences when genetic drift and mutation are acting (via the coalescent)

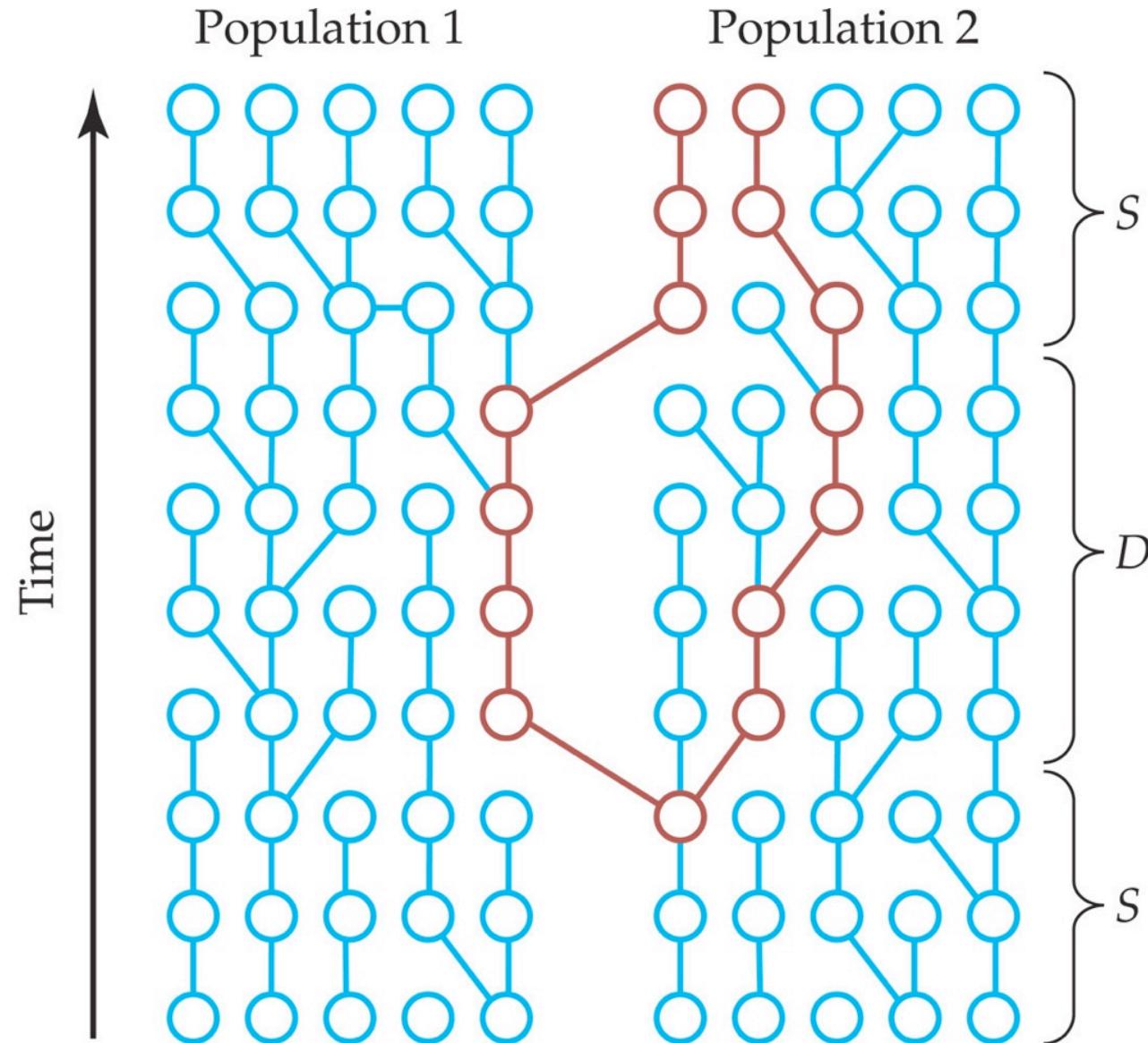
This week: the effect of subdivided populations (a.k.a. “population structure”) and migration

outline

- Population structure
- Direct methods for studying dispersal: Parentage analysis
- The effects of population subdivision: A deficiency of heterozygosity
- Measuring the deficiency of heterozygosity using F-statistics

population structure

Can be understood in a coalescent framework

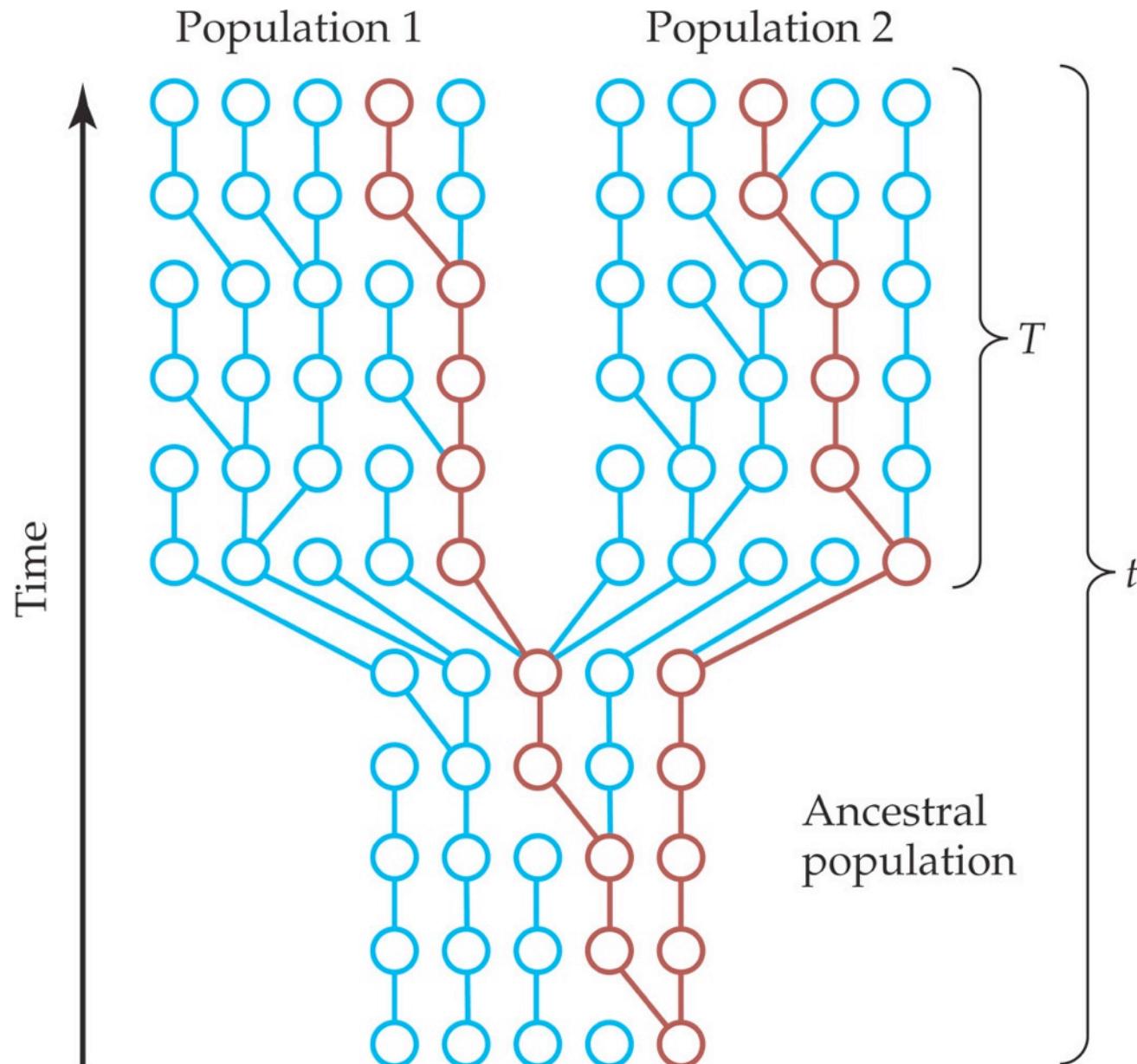


INTRODUCTION TO POPULATION GENETICS, Figure 4.4

© 2013 Sinauer Associates, Inc.

population structure

Can be understood in a coalescent framework



INTRODUCTION TO POPULATION GENETICS, Figure 4.6

© 2013 Sinauer Associates, Inc.

population structure

- A population is said to have *population structure* or *population subdivision* when matings occur within subgroups more often than expected from random mating.

population structure

- A population is said to have *population structure* or *population subdivision* when matings occur within subgroups more often than expected from random mating.
- We will talk about *sub-populations* or collections of *populations*.

population structure

- A population is said to have *population structure* or *population subdivision* when matings occur within subgroups more often than expected from random mating.
- We will talk about *sub-populations* or collections of *populations*.
- Population structure results in genetic differentiation

population structure

- A population is said to have *population structure* or *population subdivision* when matings occur within subgroups more often than expected from random mating.
- We will talk about *sub-populations* or collections of *populations*.
- Population structure results in genetic differentiation
- Two major sources of differentiation:

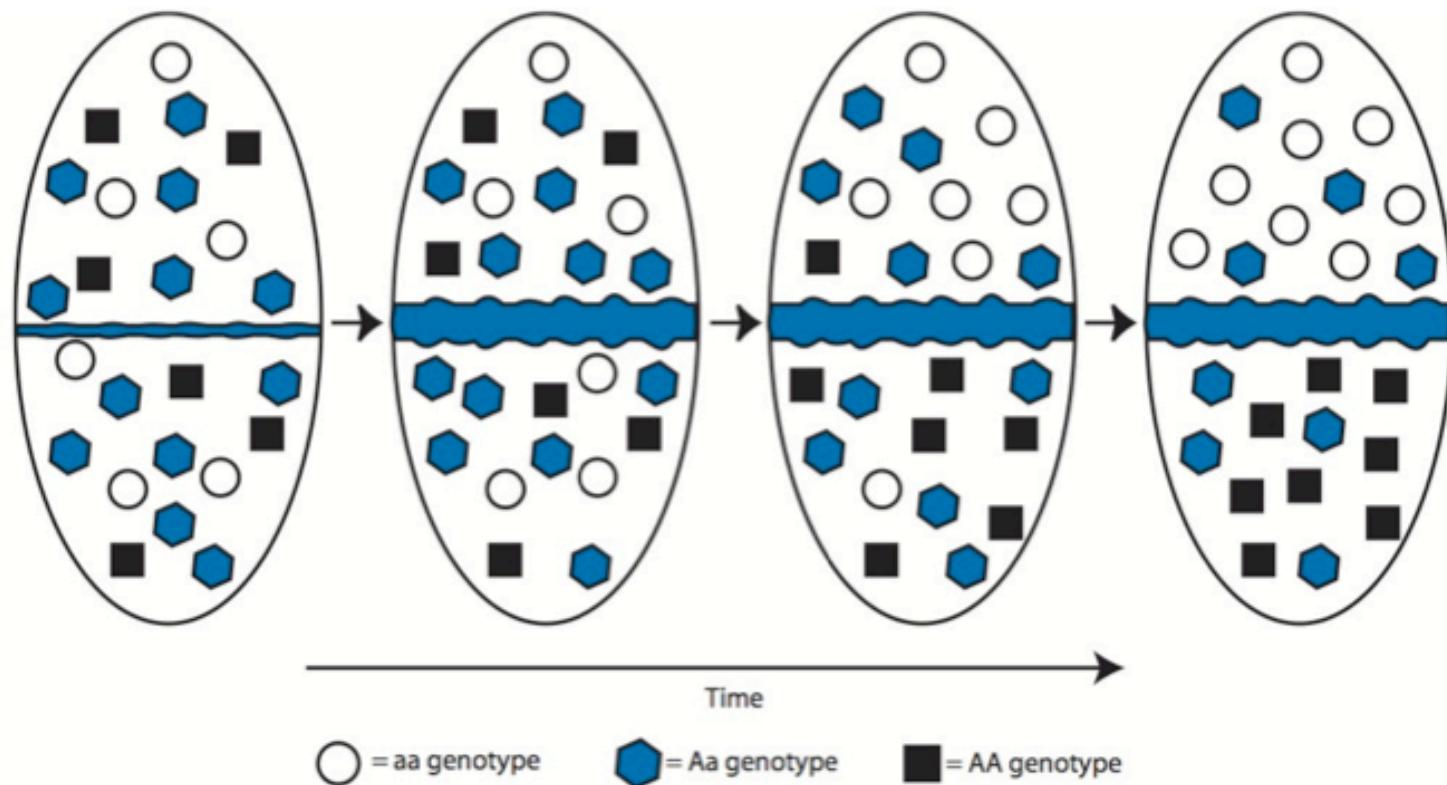
population structure

- A population is said to have *population structure* or *population subdivision* when matings occur within subgroups more often than expected from random mating.
- We will talk about *sub-populations* or collections of *populations*.
- Population structure results in genetic differentiation
- Two major sources of differentiation:
 - Selection - can act in one sub-population and not another; or favor different alleles in different sub-populations.

population structure

- A population is said to have *population structure* or *population subdivision* when matings occur within subgroups more often than expected from random mating.
- We will talk about *sub-populations* or collections of *populations*.
- Population structure results in genetic differentiation
- Two major sources of differentiation:
 - Selection - can act in one sub-population and not another; or favor different alleles in different sub-populations.
 - Drift - by chance alone, allele frequencies can become different among sub-populations.

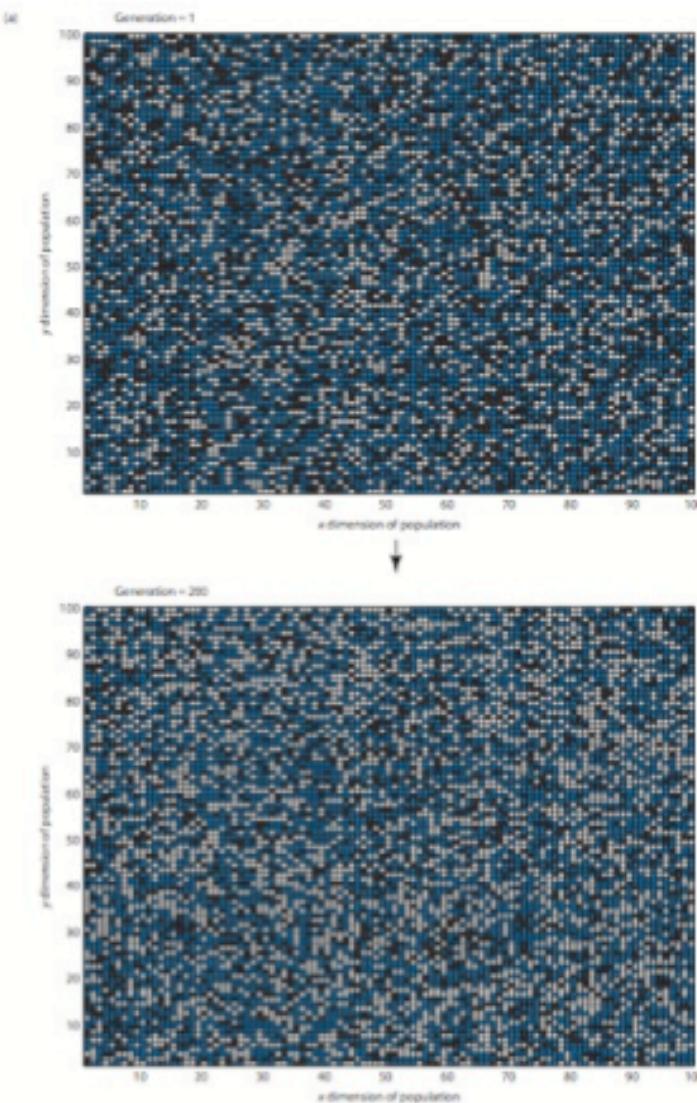
Example of population structure and divergence due to a vicariant event



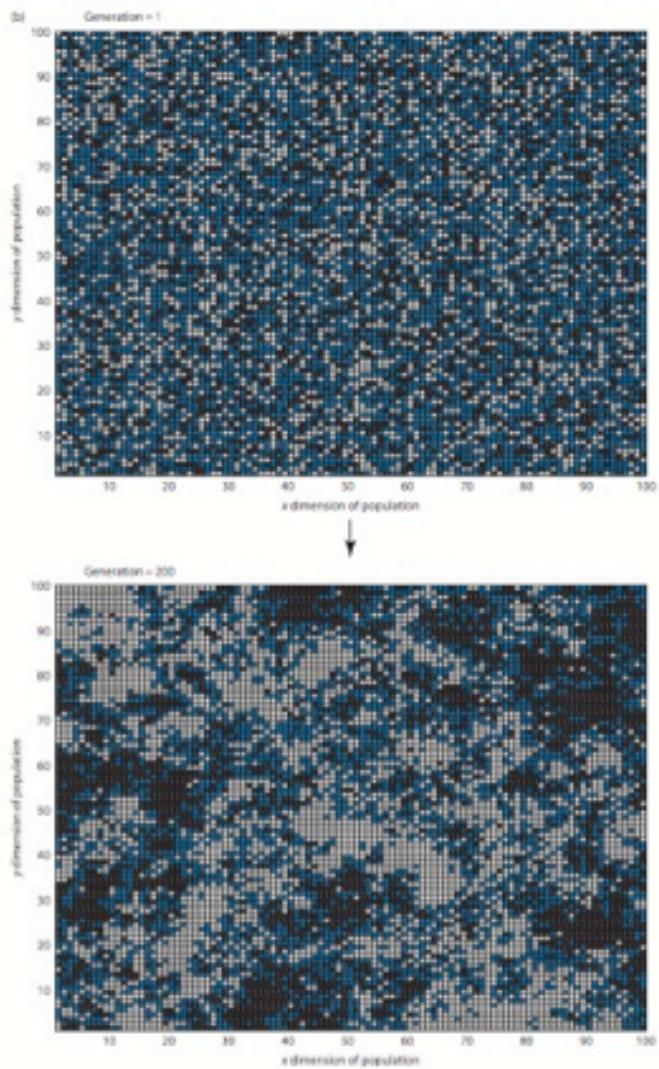
- Gene flow
- Vicariance event
- Panmixia

Example of population structure due to “isolation by distance”

Random mating



Mating only within 3×3 neighborhood



Desert wildflowers: *Linanthus parryae*



Figure 1. Blue- and white-flowered plants of *Linanthus parryae* from the Mojave Desert near Pearblossom, California, USA. D. W. Schemske.



2005

Direct measures of gene flow: Genetic marker-based parentage analysis

We can study gene flow by observing who mates with whom in a population.

Question

How can we observe who mates with whom for species where mating is not directly observable? (e.g. plants)

Direct measures of gene flow: Genetic marker-based parentage analysis

We can study gene flow by observing who mates with whom in a population.

Question

How can we observe who mates with whom for species where mating is not directly observable? (e.g. plants)

Answer: Parentage analysis

Use genetic markers to determine the parents of an individual. A typical application in trees:

Direct measures of gene flow: Genetic marker-based parentage analysis

We can study gene flow by observing who mates with whom in a population.

Question

How can we observe who mates with whom for species where mating is not directly observable? (e.g. plants)

Answer: Parentage analysis

Use genetic markers to determine the parents of an individual. A typical application in trees:

- Genotype a seed from a known tree (i.e. mother of seed is hence known).

Direct measures of gene flow: Genetic marker-based parentage analysis

We can study gene flow by observing who mates with whom in a population.

Question

How can we observe who mates with whom for species where mating is not directly observable? (e.g. plants)

Answer: Parentage analysis

Use genetic markers to determine the parents of an individual. A typical application in trees:

- Genotype a seed from a known tree (i.e. mother of seed is hence known).
- Genotype potential father trees and identify father by matching the seed genotype.

Direct measures of gene flow: Genetic marker-based parentage analysis

We can study gene flow by observing who mates with whom in a population.

Question

How can we observe who mates with whom for species where mating is not directly observable? (e.g. plants)

Answer: Parentage analysis

Use genetic markers to determine the parents of an individual. A typical application in trees:

- Genotype a seed from a known tree (i.e. mother of seed is hence known).
- Genotype potential father trees and identify father by matching the seed genotype.
- Use observed distance between father and mother to infer the distance which pollen travels.

Showing 1246 of 1246 Relatives

Sort by

Strength of Relationship

		mike hickerson Owned profile	Son 47.6% DNA shared, 22 segments
		Vivian Joiner Pierce Connected	Second to Third Cousin 1.89% DNA shared, 9 segments
		Penelope Sweet	Second to Third Cousin 1.91% DNA shared, 6 segments
		Mark McDougall Connected	Second to Third Cousin 1.50% DNA shared, 5 segments
		Jennifer Henderson	Second to Fourth Cousin 1.32% DNA shared, 5 segments
		Richard Ellerbeck Connected	Second to Fourth Cousin 1.29% DNA shared, 4 segments

Detecting population subdivision

Question

How can we easily detect that there is population subdivision?

Deficiency of observed heterozygosity due to population subdivision

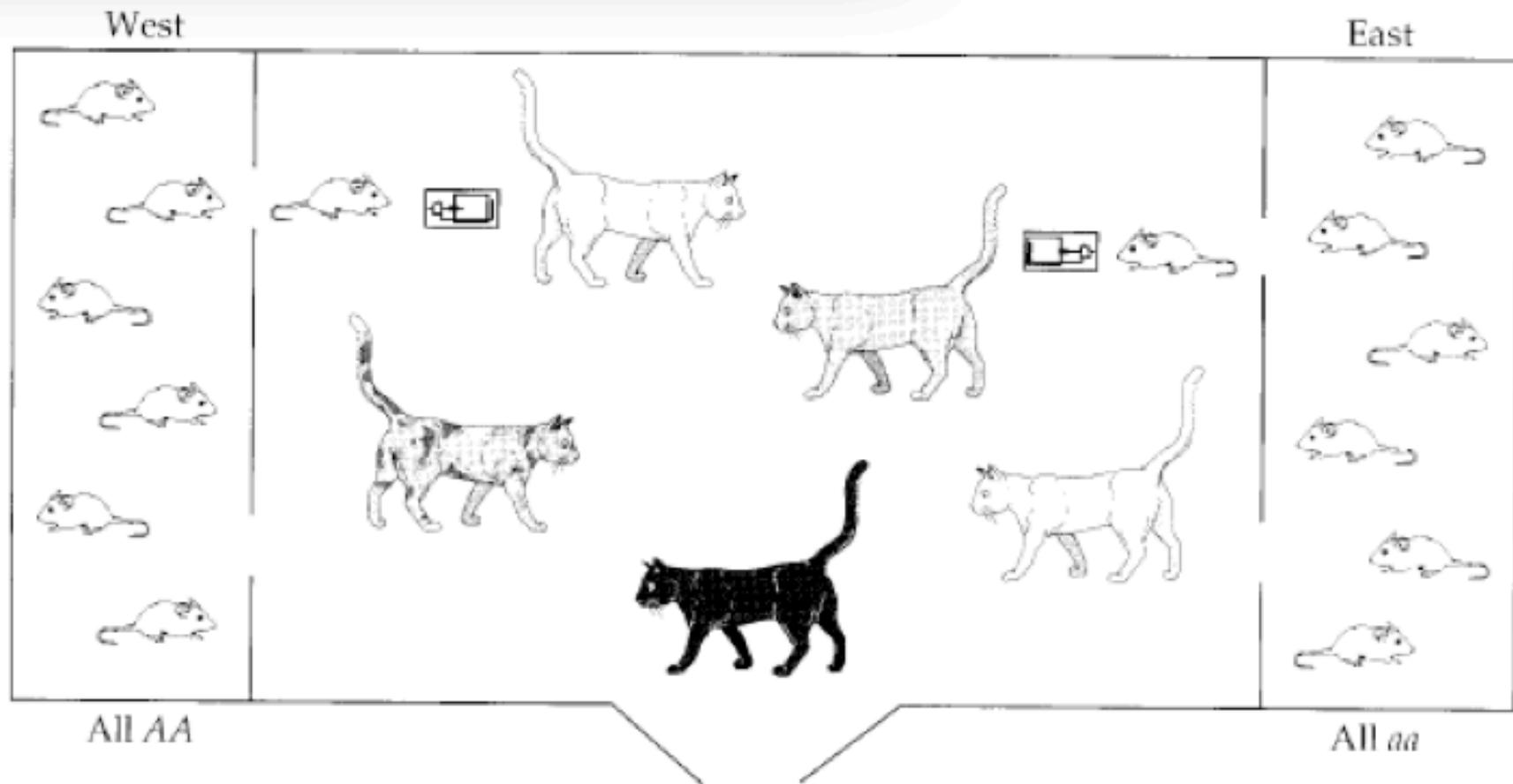


FIGURE 6.12 An extreme example of the general principle that a difference in allele frequency among subpopulations results in a deficiency of heterozygotes. The floor plan is that of a hypothetical barn. The mouse subpopulations in the east and west enclaves are completely isolated because of the cats in the middle. The west subpopulation is fixed for the *A* allele and the east subpopulation for the *a* allele. Trapping mice at random in the area patrolled by the cats would yield an overall allele frequency of $\frac{1}{2}$, but no heterozygous genotypes.

population subdivision and inbreeding

Each subpopulation is like an “extended family.” Matings between members of the same sub-population is a form of inbreeding, in the sense that their offspring will have higher probabilities of sharing alleles from a common ancestor than matings between random members of the whole population.

population subdivision and inbreeding

Each subpopulation is like an “extended family.” Matings between members of the same sub-population is a form of inbreeding, in the sense that their offspring will have higher probabilities of sharing alleles from a common ancestor than matings between random members of the whole population.

Genotype frequencies in the presence of inbreeding

$$f_{AA} = p^2(1 - F) + pF$$

$$f_{Aa} = 2pq(1 - F)$$

$$f_{aa} = q^2(1 - F) + qF$$

Just as inbreeding leads to a reduction in heterozygosity, and we can measure it using fixation indices, we can do the same with population subdivision.

example: population structure with local inbreeding



Population	n_{AA}	n_{Aa}	n_{aa}
Malibu	64	32	4
Catalina Island	49	42	9



- ① Calculate allele frequency in the two populations (p_1 and p_2 and overall \bar{p}).
- ② Calculate observed heterozygosity for each population (H_i , for $i = 1, 2$)
- ③ Calculate expected heterozygosity for each population ($2p_i(1 - p_i)$ for $i = 1, 2$)
- ④ Calculate expected heterozygosity for total population ($2\bar{p}(1 - \bar{p})$)
- ⑤ Compare observed to expected at each scale

Heterozygosities

Table 4.5 The mathematical and biological definitions of heterozygosity for three levels of population organization. In the summations, i refers to each subpopulation 1, 2, 3 . . . 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.
$H_T = 2\bar{p}\bar{q}$	The expected heterozygosity of the total population assuming random mating within subpopulations and no divergence of allele frequencies among subpopulations.

measuring population subdivision: Wright's F statistics

To measure the inbreeding effect, Sewall Wright defined the *fixation index*. The basic idea: **to measure the reduction in heterozygosity due to levels of sub-division**. The reduction from H_T to H_S :

$$(\text{Difference in } H) = H_T - H_S$$

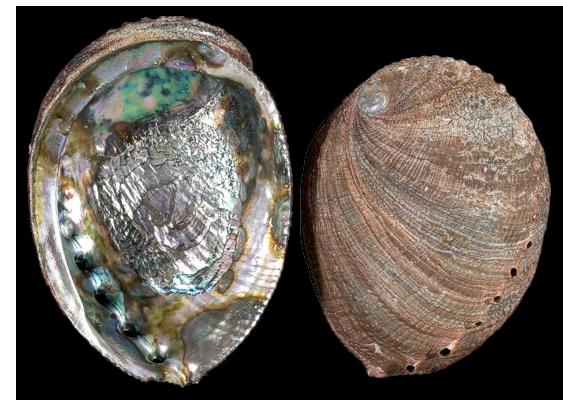
but to standardize the statistic, we compare the reduction, to its maximal value and define the statistic:

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

or in words:

$$F_{ST} = \frac{\text{Reduction in heterozygosity in subpopulation relative to total}}{\text{Its maximum value}}$$

example: population structure with local inbreeding



Population	n_{AA}	n_{Aa}	n_{aa}
Malibu	64	32	4
Catalina Island	49	42	9



- ① Calculate allele frequency in the two populations (p_1 and p_2 and overall \bar{p}).
- ② Calculate observed heterozygosity for each population (H_i , for $i = 1, 2$)
- ③ Calculate expected heterozygosity for each population ($2p_i(1 - p_i)$ for $i = 1, 2$)
- ④ Calculate expected heterozygosity for total population ($2\bar{p}(1 - \bar{p})$)
- ⑤ Compare observed to expected at each scale

measuring population subdivision: Wright's F statistics

To measure effect of local inbreeding:

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

To measure effect of population structure:

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

To measure overall reduction due to both effects:

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

measuring population subdivision: Wright's F statistics

To measure effect of local inbreeding:

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

To measure effect of population structure:

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

To measure overall reduction due to both effects:

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

Definitions of heterozygosities:

- H_I : average observed heterozygosity within subpopulations: Average of $n_{Aa}/2n$ across population
- H_S : average expected heterozygosity within subpopulations: Average of $2pq$ across population
- H_T : expected heterozygosity of total population: $2\bar{p}\bar{q}$

F -statistics

- Theoretical minimum and maximum of an F -statistic: 0 and 1.



$$F \sim 0$$

st



$$F_{\text{st}} > 0.5$$

F -statistics

- Theoretical minimum and maximum of an F -statistic: 0 and 1.
- Specialized estimators exist for small sample sizes (Weir and Cockerham 1984).

F -statistics

- Theoretical minimum and maximum of an F -statistic: 0 and 1.
- Specialized estimators exist for small sample sizes (Weir and Cockerham 1984).
- Rules of thumb for F_{ST} :
 - 0-0.05: “Little” genetic differentiation

F -statistics

- Theoretical minimum and maximum of an F -statistic: 0 and 1.
- Specialized estimators exist for small sample sizes (Weir and Cockerham 1984).
- Rules of thumb for F_{ST} :
 - 0-0.05: “Little” genetic differentiation
 - 0.05-0.15: “Moderate” genetic differentiation

F -statistics

- Theoretical minimum and maximum of an F -statistic: 0 and 1.
- Specialized estimators exist for small sample sizes (Weir and Cockerham 1984).
- Rules of thumb for F_{ST} :
 - 0-0.05: “Little” genetic differentiation
 - 0.05-0.15: “Moderate” genetic differentiation
 - 0.15-0.25: “Great” genetic differentiation

F -statistics

- Theoretical minimum and maximum of an F -statistic: 0 and 1.
- Specialized estimators exist for small sample sizes (Weir and Cockerham 1984).
- Rules of thumb for F_{ST} :
 - 0-0.05: “Little” genetic differentiation
 - 0.05-0.15: “Moderate” genetic differentiation
 - 0.15-0.25: “Great” genetic differentiation
 - > 0.25 : “Very great”

F-statistics

- Theoretical minimum and maximum of an *F*-statistic: 0 and 1.
- Specialized estimators exist for small sample sizes (Weir and Cockerham 1984).
- Rules of thumb for F_{ST} :
 - 0-0.05: “Little” genetic differentiation
 - 0.05-0.15: “Moderate” genetic differentiation
 - 0.15-0.25: “Great” genetic differentiation
 - > 0.25: “Very great”
- But these rules of thumb should be taken with a grain of salt: The true measure of differentiation should depend on your question/application.

additional F -statistics tidbits

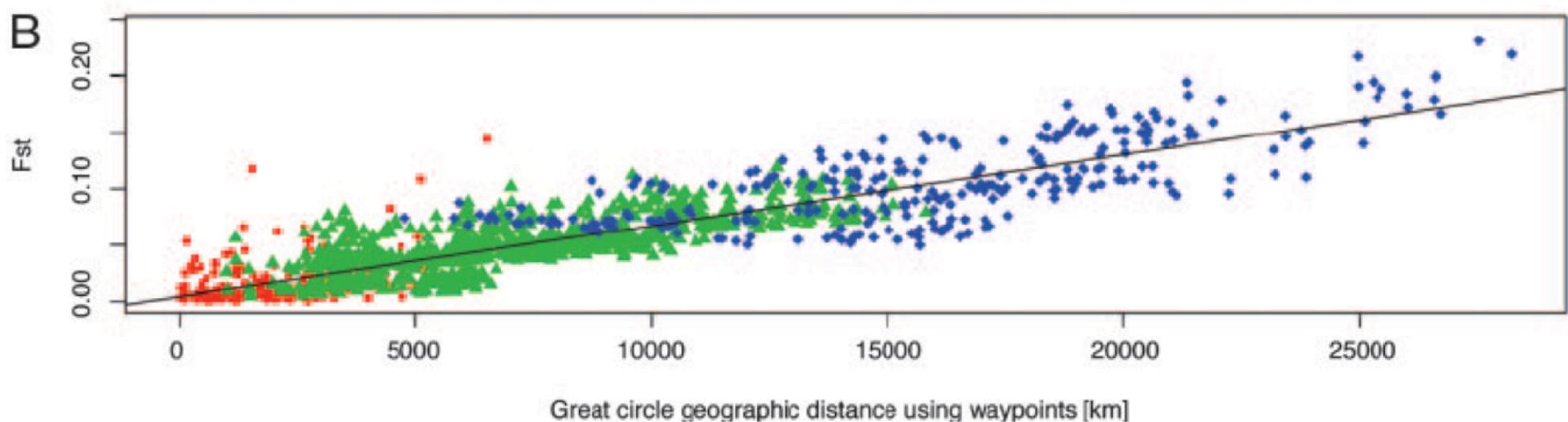
- F_{ST} can be written in terms of the *variance* and *mean* in allele frequencies across populations (σ^2 and \bar{p} , \bar{q}):

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

where $\sigma^2 = \sum_{k=1}^K (p_i - \bar{p})^2$

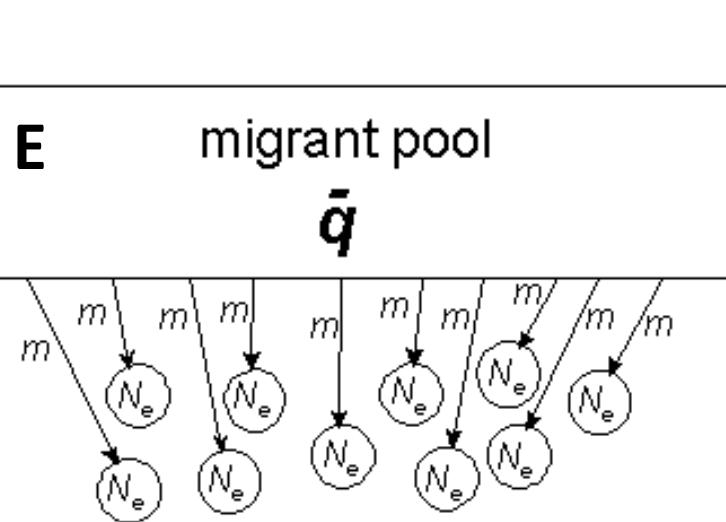
human

If we sample[^]sub-populations from across the world, $F_{ST} \approx 0.05$. This is “moderate” genetic differentiation.

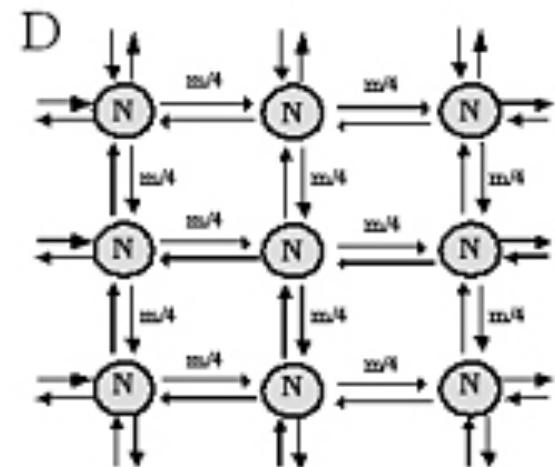
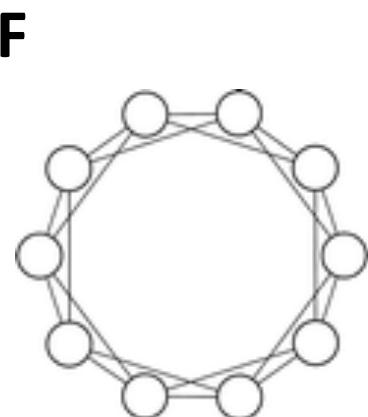
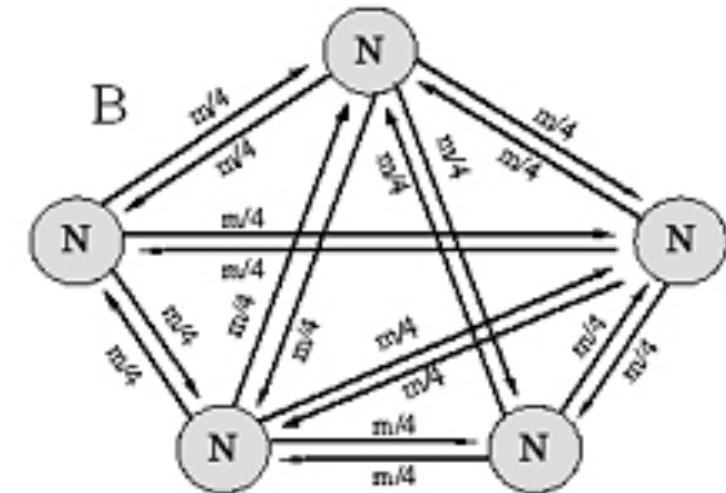
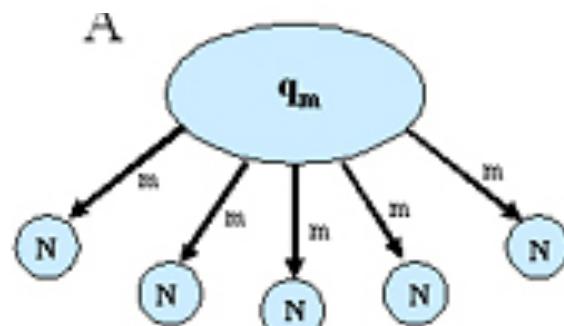


Plot of F_{ST} calculated for two sub-populations as function of the geographic distance (measured via waypoints) between them.

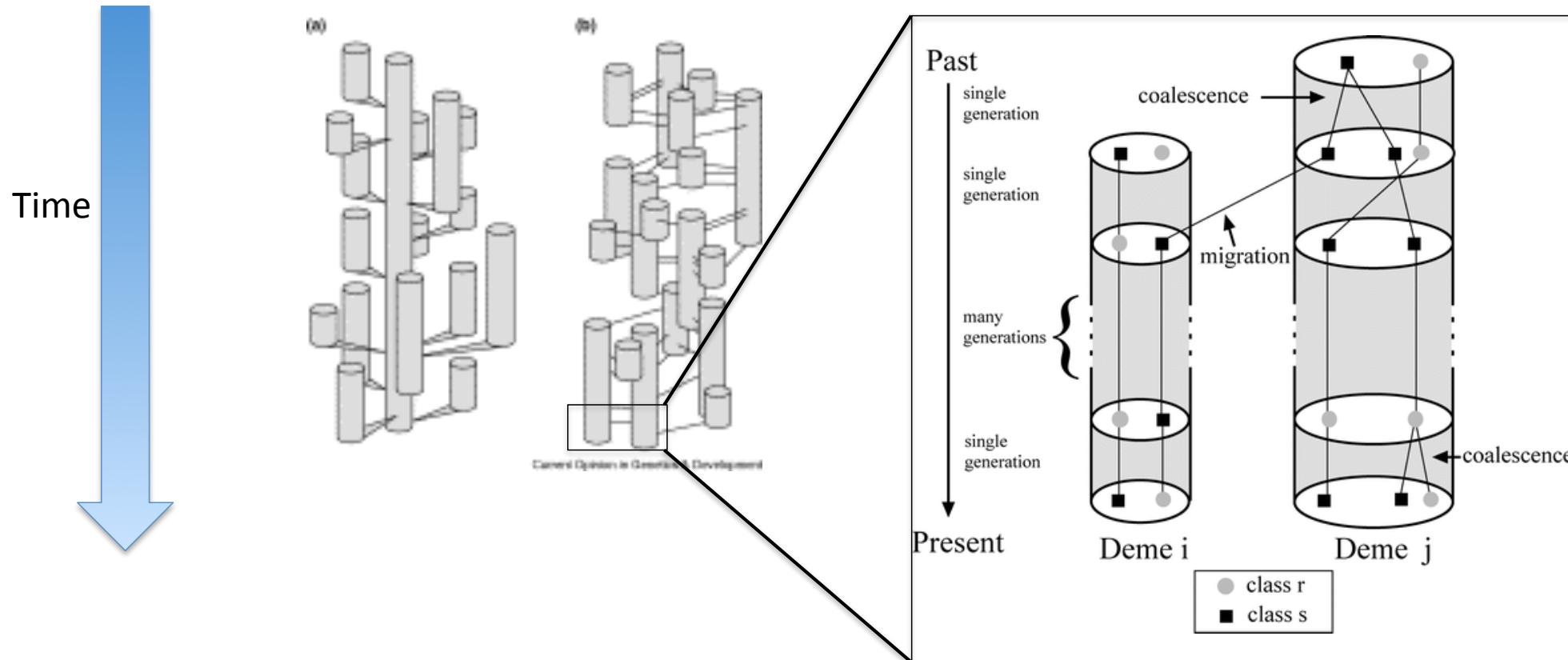
Models of migration



- A. Island – Continent model**
- B & E. Island**
- C. Stepping stone**
- D. Lattice**
- F. Ring Species**

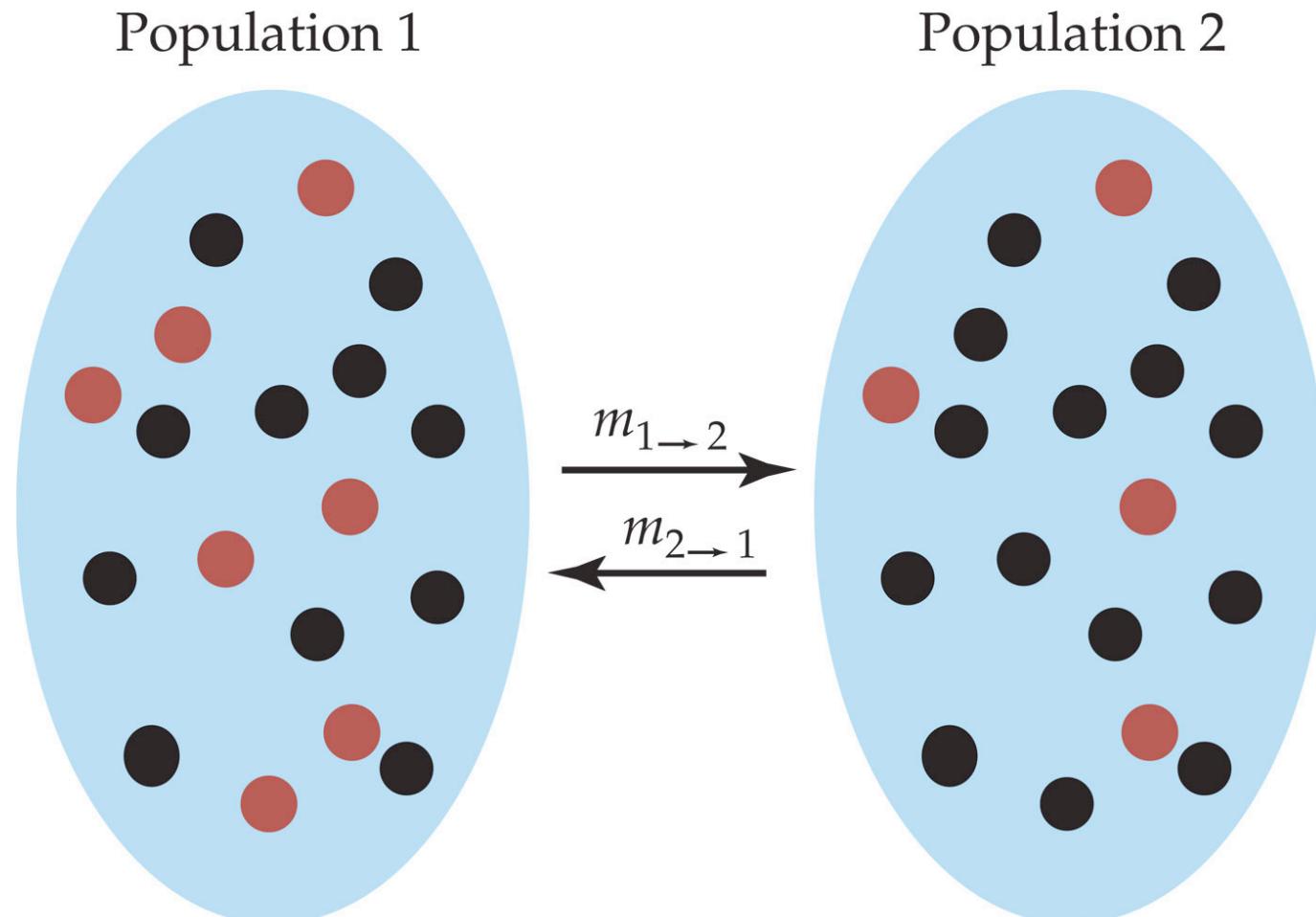


Meta-populations (many small populations appearing and going extinct)



More populations = higher global N_e = higher genetic diversity

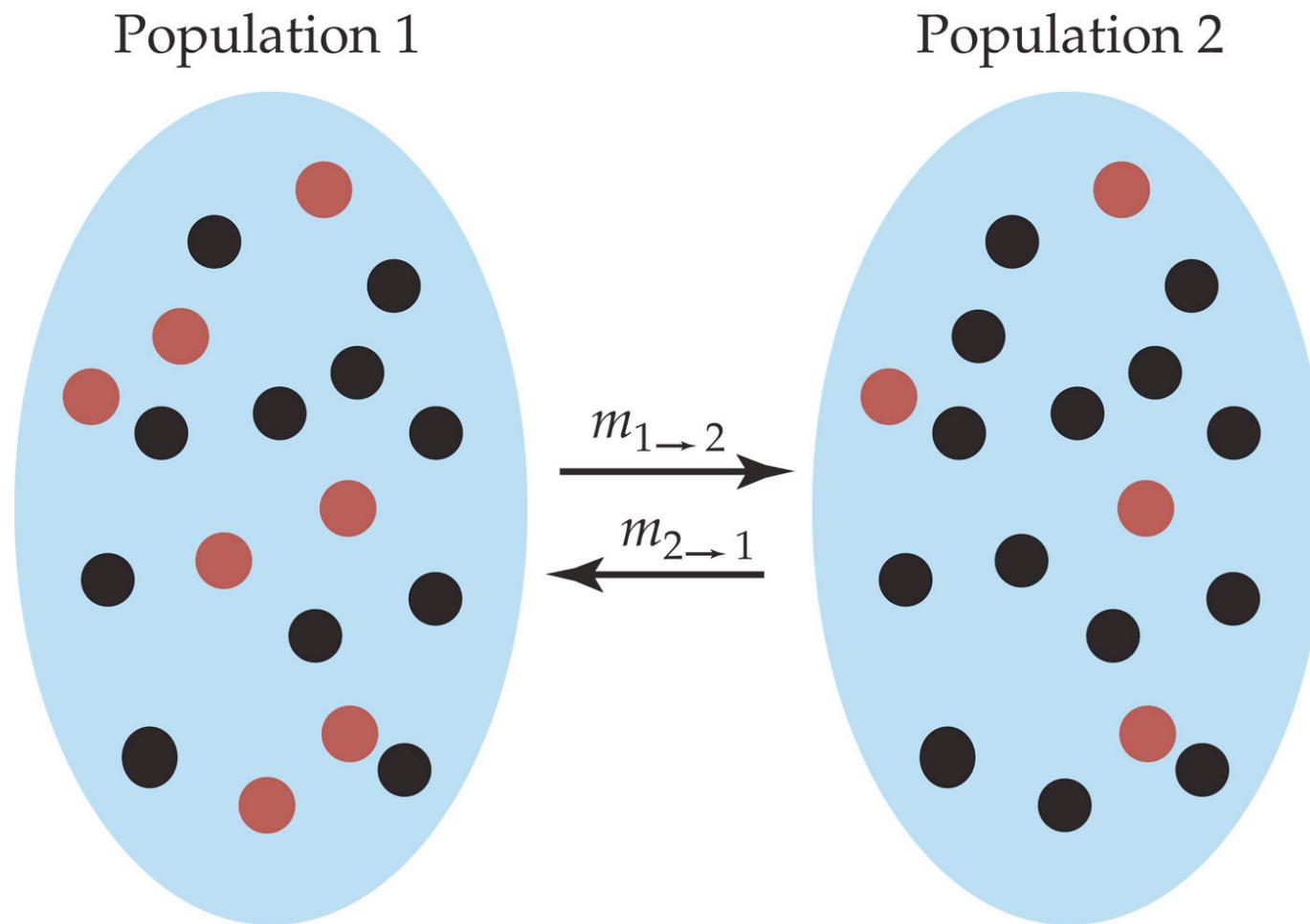
Models of migration - two Island model



INTRODUCTION TO POPULATION GENETICS, Figure 4.2
© 2013 Sinauer Associates, Inc.

m is the per generation probability of migration

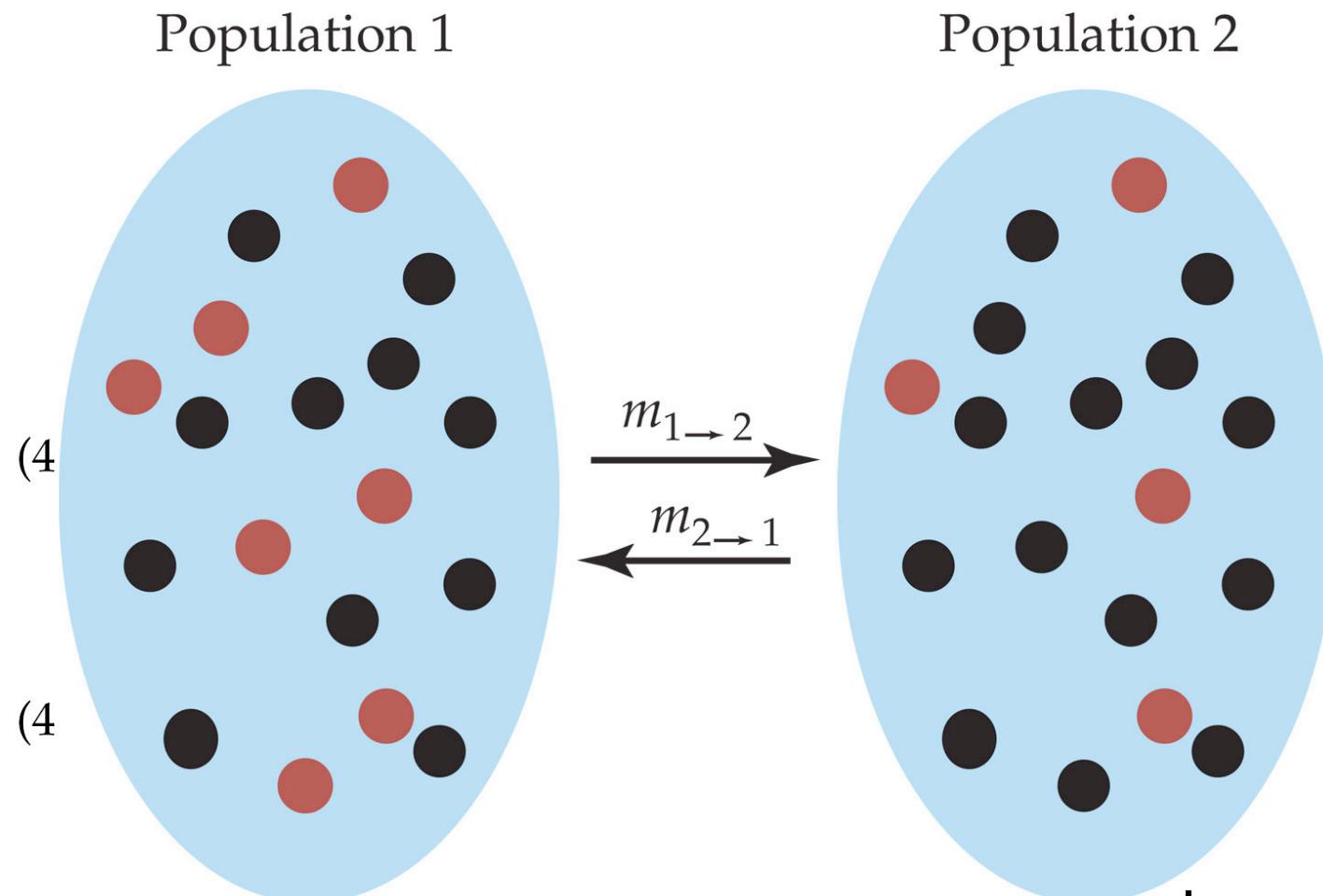
Models of migration - two Island model



INTRODUCTION TO POPULATION GENETICS, Figure 4.2
© 2013 Sinauer Associates, Inc.

What is the frequency of allele A in pops 1 & 2 at the next time step $t + 1$?

Models of migration - two Island model



INTRODUCTION TO POPULATION GENETICS, Figure 4.2
© 2013 Sinauer Associates, Inc.

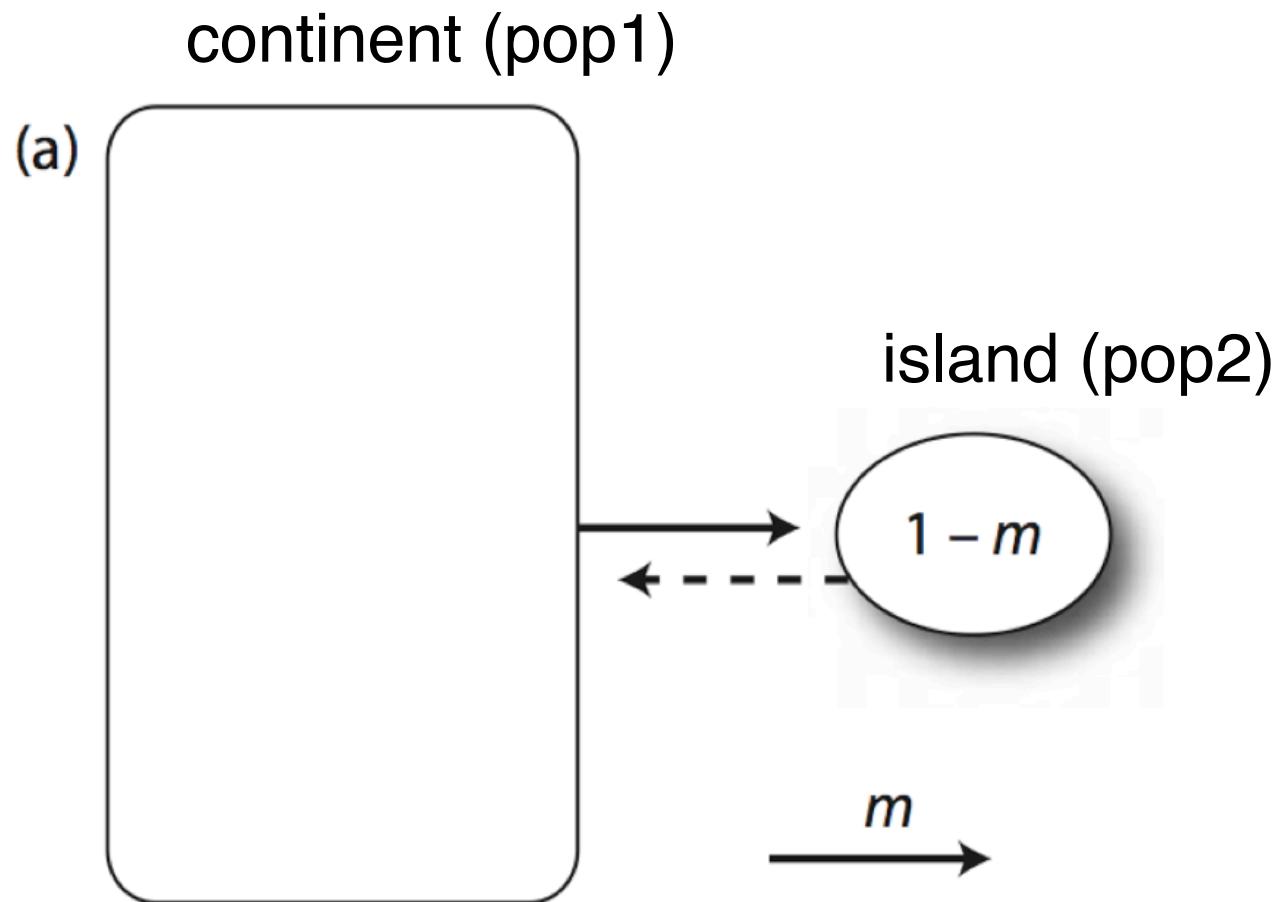
$$(4.7) \quad E[f_{A1}(t+1)] = (1 - m_{2 \rightarrow 1})f_{A1}(t) + m_{2 \rightarrow 1} f_{A2}(t)$$

depends on initial freq
and the migration rates

imagine cranking m to 1.0

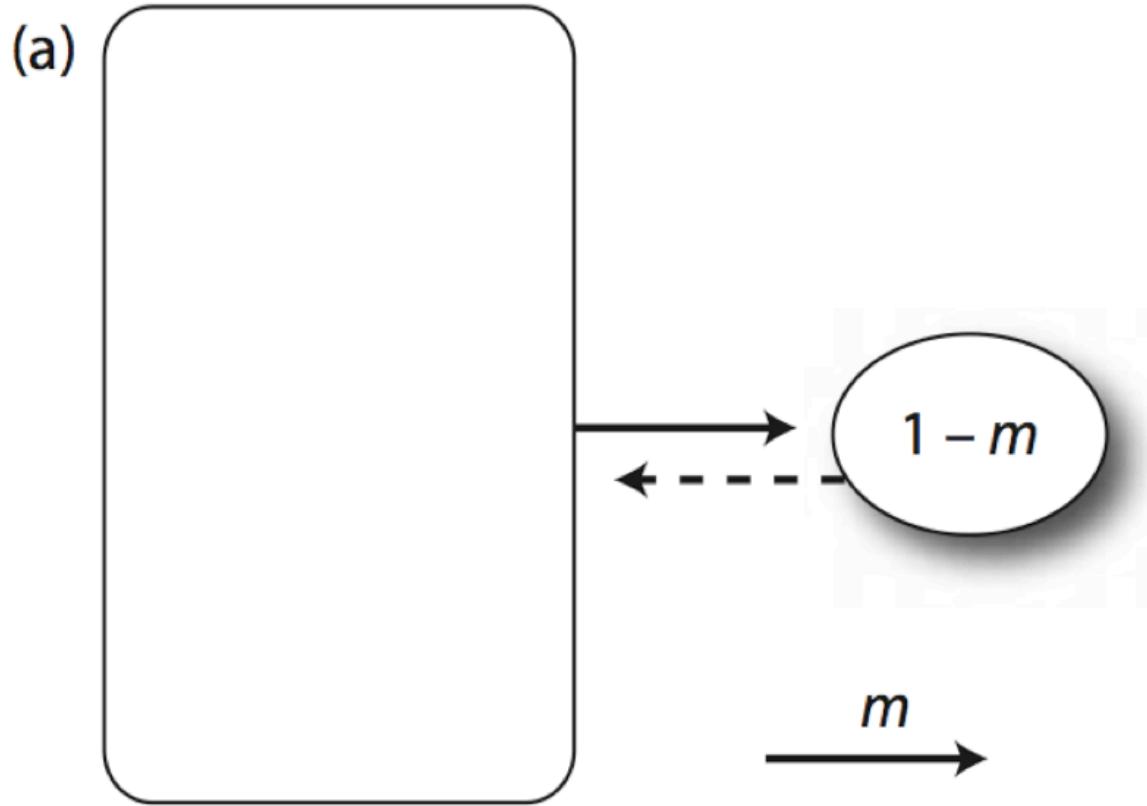
$$(4.8) \quad E[f_{A2}(t+1)] = (1 - m_{1 \rightarrow 2})f_{A2}(t) + m_{1 \rightarrow 2} f_{A1}(t)$$

Models of migration - Continent /Island model



NOTE: Continent is so large - the impact of migration from the island is negligible.

m is the per generation probability of migration



$$(4.7) \quad E[f_{A1}(t+1)] = (1 - \cancel{m_{2 \rightarrow 1}})f_{A1}(t) + \cancel{m_{2 \rightarrow 1} f_{A2}(t)}$$

just becomes a single pop drift model

$$(4.8) \quad E[f_{A2}(t+1)] = (1 - m_{1 \rightarrow 2})f_{A2}(t) + m_{1 \rightarrow 2} f_{A1}(t)$$

Fst in a model with genetic drift

- Imagine a large number of islands that each start with an initial allele frequency p_0 and then undergo genetic drift in complete isolation of one another.

F_{ST} in a model with genetic drift

- Imagine a large number of islands that each start with an initial allele frequency p_0 and then undergo genetic drift in complete isolation of one another.
- Each island is losing heterozygosity due to drift, but the collection of islands maintains an average allele frequency $\bar{p} = p_0$ and $H_T = 2\bar{p}\bar{q}$.

F_{ST} in a model with genetic drift

- Imagine a large number of islands that each start with an initial allele frequency p_0 and then undergo genetic drift in complete isolation of one another.
- Each island is losing heterozygosity due to drift, but the collection of islands maintains an average allele frequency $\bar{p} = p_0$ and $H_T = 2\bar{p}\bar{q}$.
- H_S will decay as

$$H_0 e^{-\frac{t}{2N}}$$

F_{ST} in a model with genetic drift

- Imagine a large number of islands that each start with an initial allele frequency p_0 and then undergo genetic drift in complete isolation of one another.
- Each island is losing heterozygosity due to drift, but the collection of islands maintains an average allele frequency $\bar{p} = p_0$ and $H_T = 2\bar{p}\bar{q}$.
- H_S will decay as

$$H_0 e^{-\frac{t}{2N}}$$

- H_T stays constant. Note also that H_S begins with the value H_T (i.e $H_0 = H_T$) So

$$F_{ST}(t) = \frac{H_T - H_T e^{-\frac{t}{2N}}}{H_T} = 1 - e^{-\frac{t}{2N}}$$

(ie approach 1 as time marches on with total isolation)

Drift vs. Migration

- In models with genetic drift and no migration (i.e. pure isolation), heterozygosity decays within populations, and allele frequencies disperse due to drift. F_{ST} increases to one.

Drift vs. Migration

- In models with genetic drift and no migration (i.e. pure isolation), heterozygosity decays within populations, and allele frequencies disperse due to drift. F_{ST} increases to one.
- In models with migration, allele frequencies converge to a single value. Heterozygosity is maintained and equals the total heterozygosity. F_{ST} should decrease to zero.

Drift vs. Migration

- In models with genetic drift and no migration (i.e. pure isolation), heterozygosity decays within populations, and allele frequencies disperse due to drift. F_{ST} increases to one.
- In models with migration, allele frequencies converge to a single value. Heterozygosity is maintained and equals the total heterozygosity. F_{ST} should decrease to zero.
- In models with both drift and migration, overtime a *migration-drift* equilibrium arises which dictates the variance in allele frequencies.

drift migration equilibrium

under a 2 population model

$$F_{st} = 1/(1 + 8M) \quad (M = Nm)$$

under a model with “many” populations (ie island model)

$$F_{st} = 1/(1 + 4Nm)$$

- Under the symmetric island model, at equilibrium:

$$F_{ST} = \frac{1}{1 + 4Nm}$$

- Nm = the number of migrant individuals per generation
- The equilibrium F_{ST} quickly falls as a function of Nm :

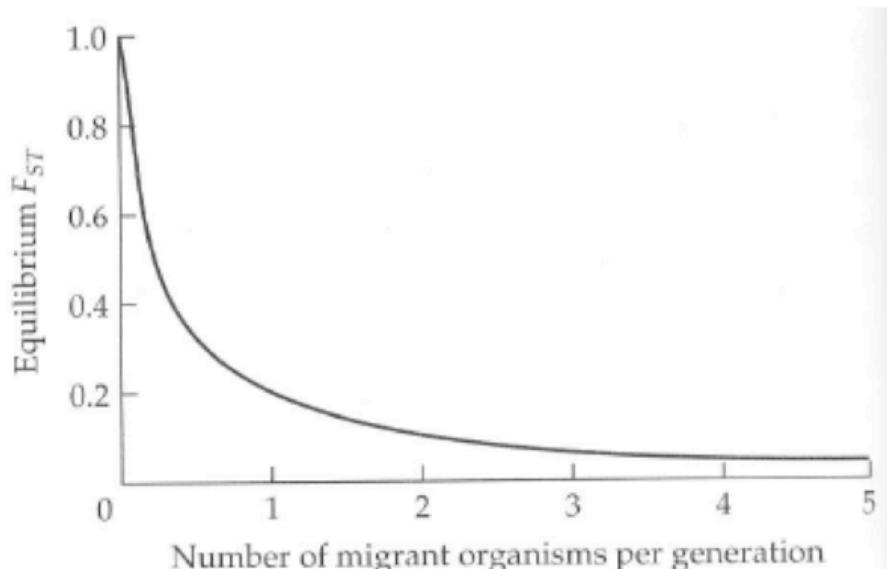


FIGURE 6.20 Decrease in the fixation index F_{ST} among subpopulations at equilibrium in the island model of migration. The curve is that in Equation 6.23, giving \hat{F} as a function of Nm . In the island model, Nm is the number of migrant organisms that come into each subpopulation in each generation.

- Under the symmetric island model, at equilibrium:

$$F_{ST} = 1/(1 + 8M)$$

$$F_{ST} = \frac{1}{1 + 4Nm}$$

- Nm = the number of migrant individuals per generation
- The equilibrium F_{ST} quickly falls as a function of Nm :

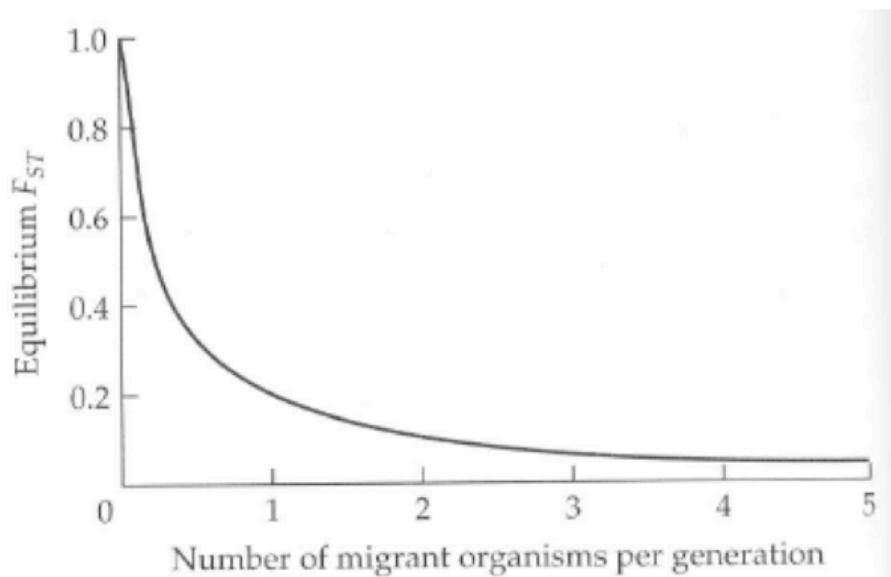
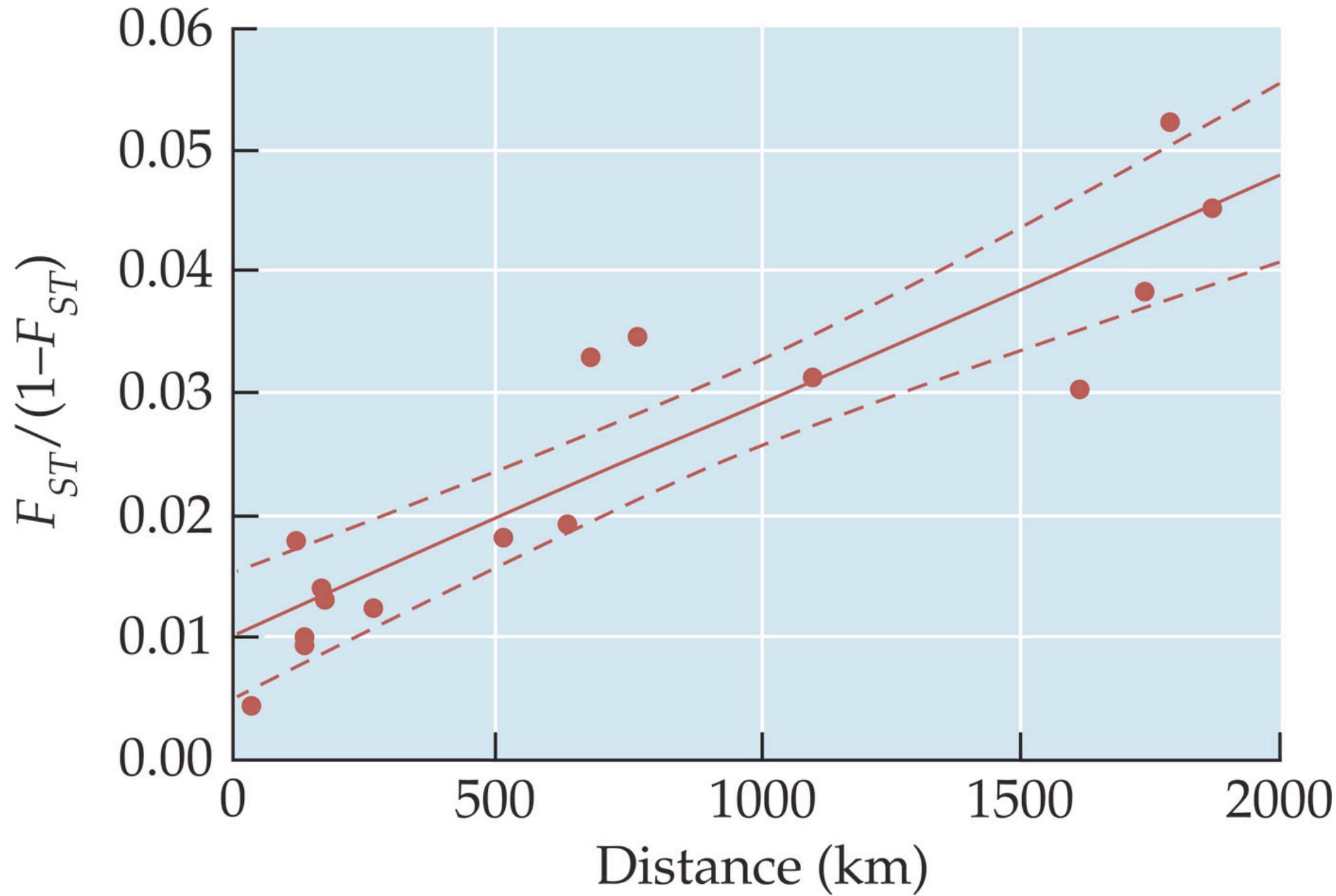


FIGURE 6.20 Decrease in the fixation index F_{ST} among subpopulations at equilibrium in the island model of migration. The curve is that in Equation 6.23, giving \hat{F} as a function of Nm . In the island model, Nm is the number of migrant organisms that come into each subpopulation in each generation.

Note

Relatively little migration is enough to keep populations from diverging greatly. $Nm > 2$ per generation implies $F_{ST} < 0.11$.

Figure 4.7 The relationship between $F_{ST}/(1 - F_{ST})$ and geographical distance (km) for California sea lion populations



indirect estimates of Nm

Table 4.9 Estimates of the fixation index among subpopulations (\hat{F}_{ST}) for diverse species based on molecular genetic marker data for nuclear loci. Different estimators were employed depending on the type of genetic marker and study design. Each \hat{F}_{ST} was used to determine the effective number of migrants ($\widehat{N_e m}$) that would produce an identical level of population structure under the assumptions of the infinite island model according to equation 4.63.

Species	\hat{F}_{ST}	$\widehat{N_e m}$	Reference
Amphibians			
<i>Alytes muletensis</i> (Mallorcan midwife toad)	0.12–0.53	1.8–0.2	Kraaijeveld-Smit et al. 2005
Birds			
<i>Gallus gallus</i> (broiler chicken breed)	0.19	1.0	Emara et al. 2002
Mammals			
<i>Capreolus capreolus</i> (roe deer)	0.097–0.146	2.2–1.4	Wang and Schreiber 2001
<i>Homo sapiens</i> (human)	0.03–0.05	7.8–4.6	Rosenberg et al. 2002
<i>Microtus arvalis</i> (common vole)	0.17	1.2	Heckel et al. 2005
Plants			
<i>Arabidopsis thaliana</i> (mouse-ear cress)	0.643	0.1	Bergelson et al. 1998
<i>Oryza officinalis</i> (wild rice)	0.44	0.3	Gao 2005
<i>Phlox drummondii</i> (annual phlox)	0.17	1.2	Levin 1977
<i>Prunus armeniaca</i> (apricot)	0.32	0.5	Romero et al. 2003
Fish			
<i>Morone saxatilis</i> (striped bass)	0.002	11.8	Brown et al. 2005
<i>Sparisoma viride</i> (stoplight parrotfish)	0.019	12.4	Geertjes et al. 2004
Insects			
<i>Drosophila melanogaster</i> (fruit fly)	0.112	2.0	Singh and Rhomberg 1987
<i>Glossina pallidipes</i> (tsetse fly)	0.18	1.1	Ouma et al. 2005
<i>Heliconius charithonia</i> (butterfly)	0.003	79.8	Kronforst and Flemming 2001
Corals			
<i>Seriatopora hystrix</i>	0.089–0.136	2.6–1.6	Maier et al. 200

indirect estimates of Nm

Table 4.9 Estimates of the fixation index among subpopulations (\hat{F}_{ST}) for diverse species based on molecular genetic marker data for nuclear loci. Different estimators were employed depending on the type of genetic marker and study design. Each \hat{F}_{ST} was used to determine the effective number of migrants ($\widehat{N_e m}$) that would produce an identical level of population structure under the assumptions of the infinite island model according to equation 4.63.

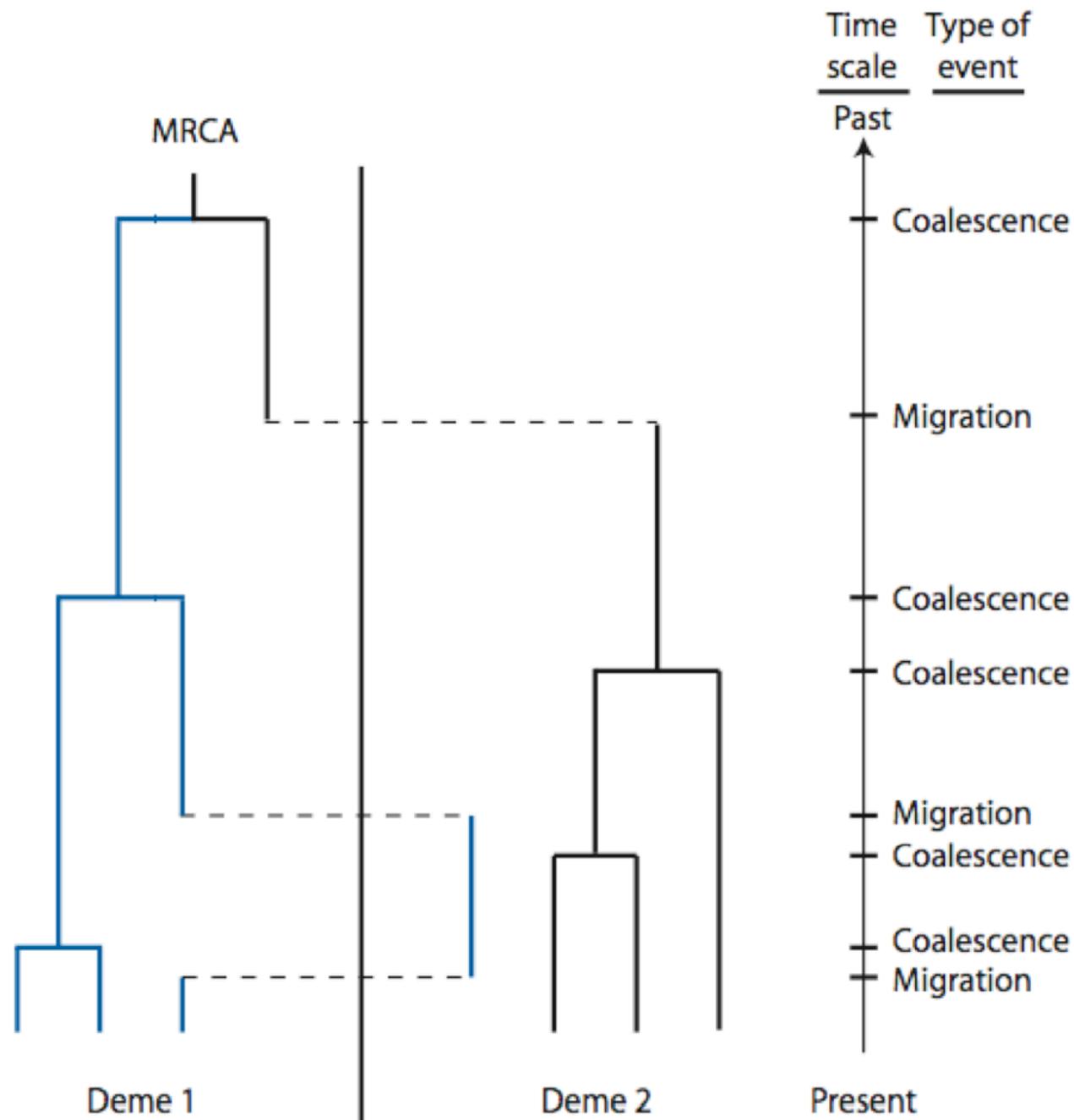
Species	\hat{F}_{ST}	$\widehat{N_e m}$	Reference
Amphibians			
<i>Alytes muletensis</i> (Mallorcan midwife toad)	0.12–0.53	1.8–0.2	Kraaijeveld-Smit et al. 2005
Birds			
<i>Gallus gallus</i> (broiler chicken breed)	0.19	1.0	Emara et al. 2002
Mammals			
<i>Capreolus capreolus</i> (roe deer)	0.097–0.146	2.2–1.4	Wang and Schreiber 2001
<i>Homo sapiens</i> (human)	0.03–0.05	7.8–4.6	Rosenberg et al. 2002
<i>Microtus arvalis</i> (common vole)	0.17	1.2	Heckel et al. 2005

Note:

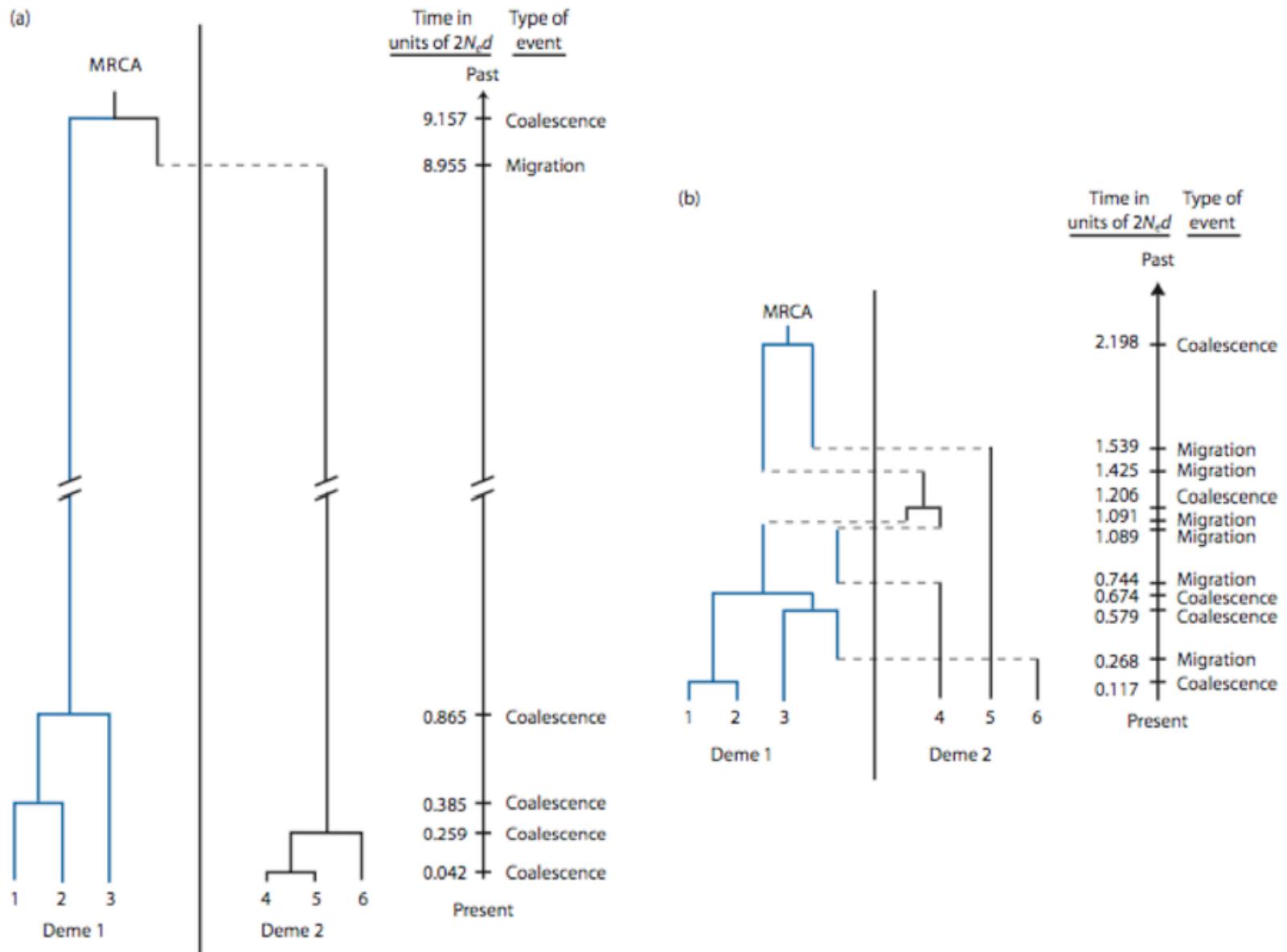
estimates should be taken with a grain of salt because few species conform to this island model. More recent approaches are usually based on coalescent-based estimators of migration that account for historical changes in important parameters

Corals	\hat{F}_{ST}	$\widehat{N_e m}$	Reference
<i>Seriatopora hystrix</i>	0.089–0.136	2.6–1.6	Maier et al. 200

Coalescent with population structure

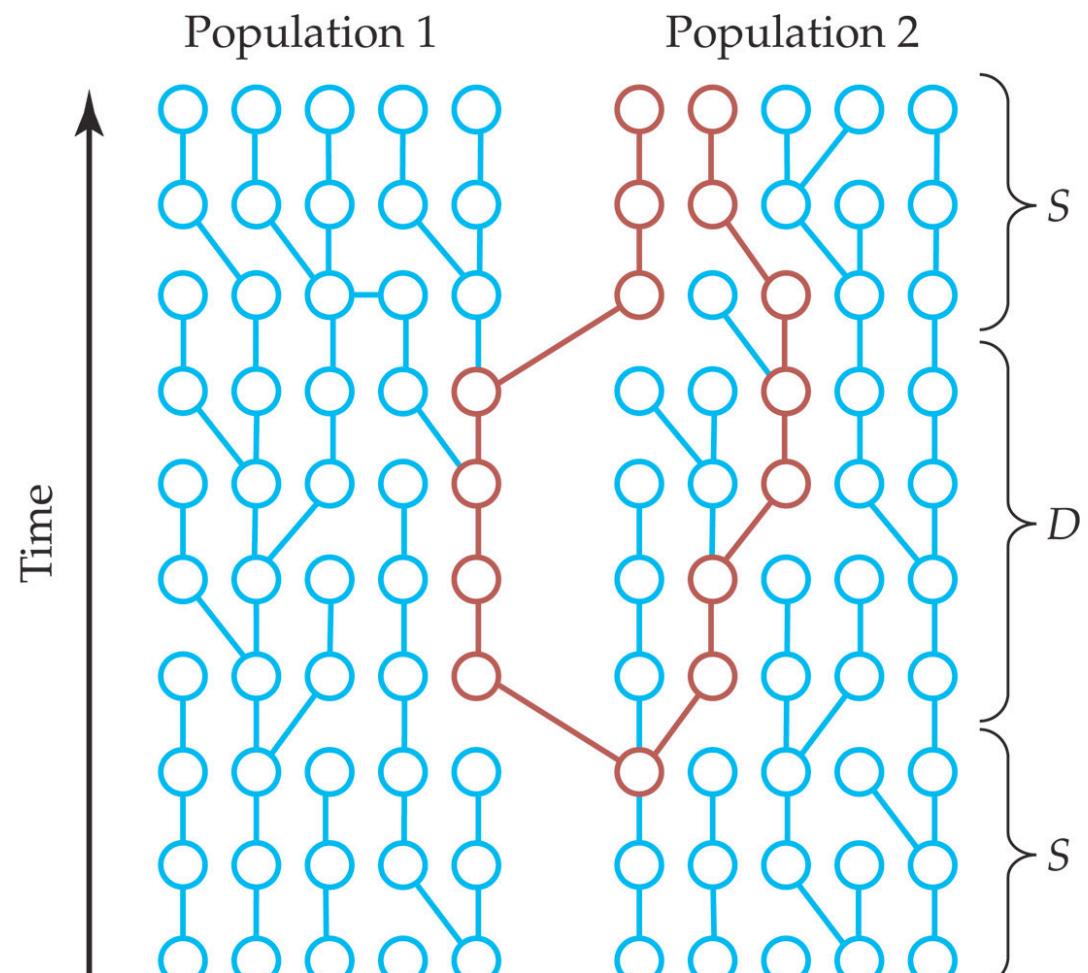


Coalescent with population structure



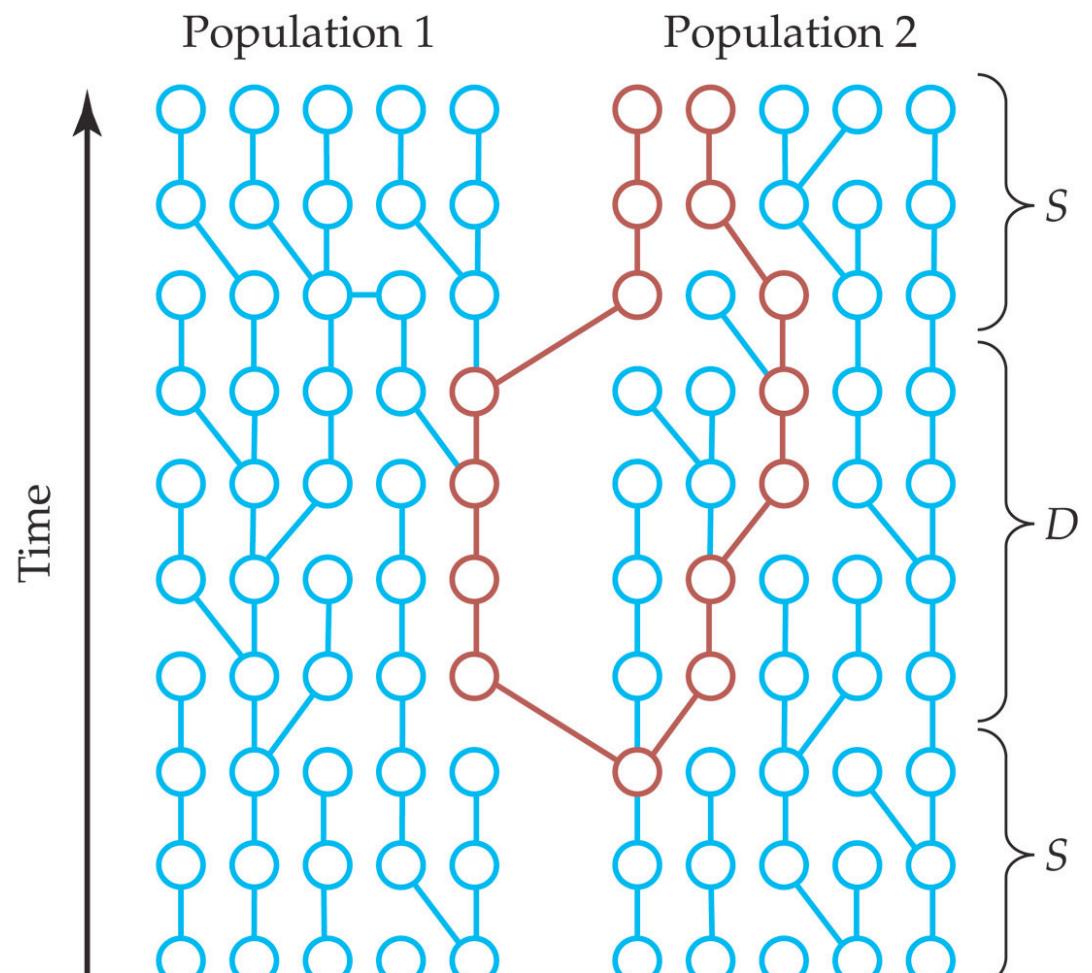
(a) Low migration rate = long internal branches; (b) high migration rate = similar to single population genealogy

$$(4.9) \quad \Pr(\text{individual was not a migrant last generation}) =$$



INTRODUCTION TO POPULATION GENETICS, Figure 4.4
© 2013 Sinauer Associates, Inc.

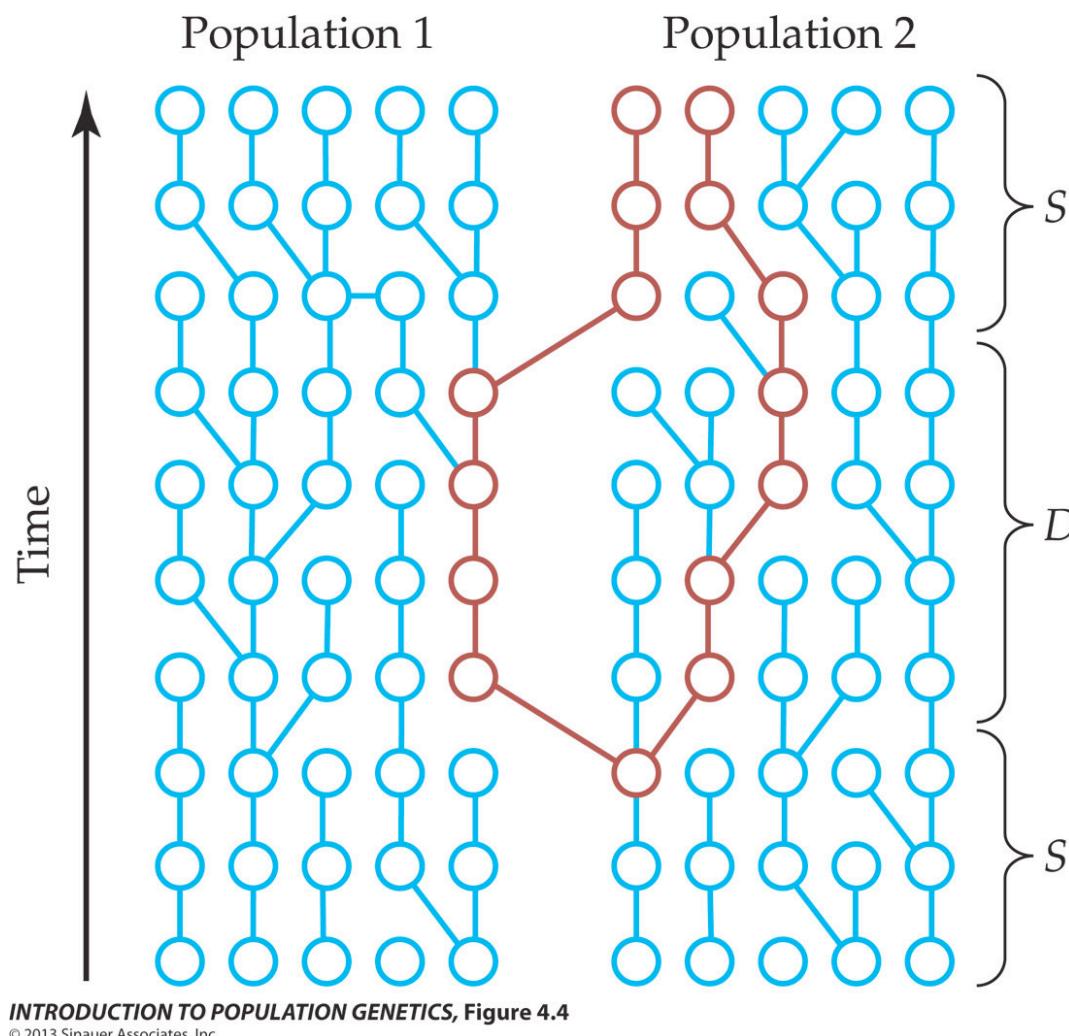
$$(4.9) \quad \Pr(\text{individual was not a migrant last generation}) = 1 - m$$



INTRODUCTION TO POPULATION GENETICS, Figure 4.4
© 2013 Sinauer Associates, Inc.

$$(4.9) \quad \Pr(\text{individual was not a migrant last generation}) = 1 - m$$

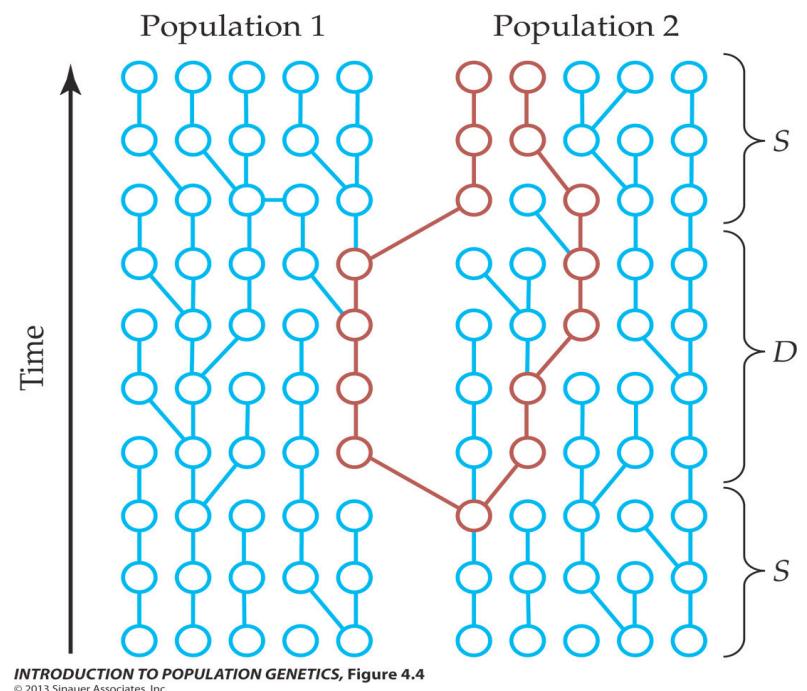
$$(4.10) \quad \Pr(\text{individual was not a migrant in the past } r \text{ generations}) = (1 - m)^r$$



$$(4.9) \quad \Pr(\text{individual was not a migrant last generation}) = 1 - m$$

$$(4.10) \quad \Pr(\text{individual was not a migrant in the past } r \text{ generations}) = (1 - m)^r$$

setting time to be in units of $2N$ generations ($r = 2Nt$), $N \rightarrow \infty$
and $M=2Nm$

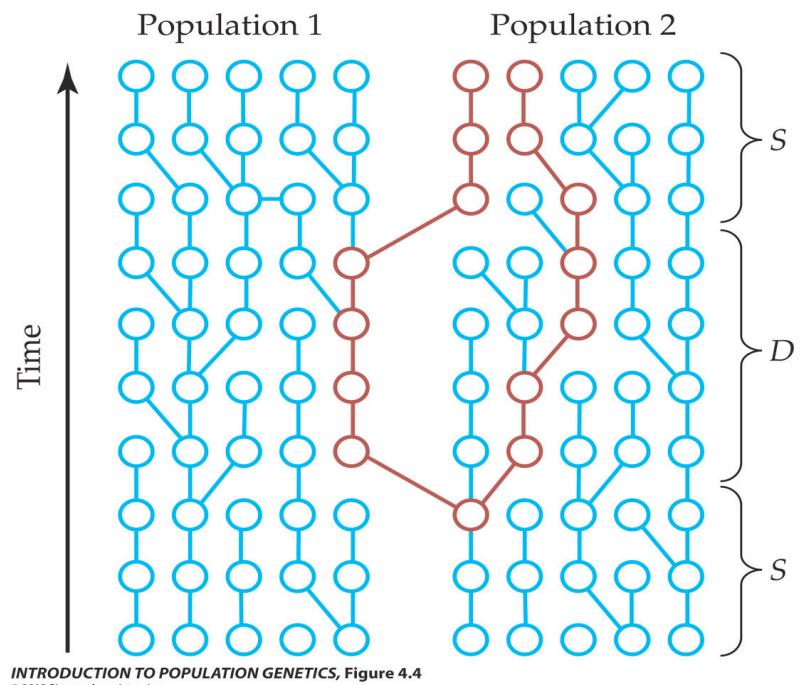


$$(4.9) \quad \Pr(\text{individual was not a migrant last generation}) = 1 - m$$

$$(4.10) \quad \Pr(\text{individual was not a migrant in the past } r \text{ generations}) = (1 - m)^r$$

setting time to be in units of $2N$ generations ($r = 2Nt$), $N \rightarrow \infty$
and $M=2Nm$ (ie $m = M/2N$), so

$$(1 - m)^r = (1 - (M/2N))^{2Nt}$$



$$(4.9) \quad \Pr(\text{individual was not a migrant last generation}) = 1 - m$$

$$(4.10) \quad \Pr(\text{individual was not a migrant in the past } r \text{ generations}) = (1 - m)^r$$

setting time to be in units of $2N$ generations ($r = 2Nt$), $N \rightarrow \infty$
and $M=2Nm$ (ie $m = M/2N$), so

$$(1 - m)^r = (1 - (M/2N))^{2Nt} - > e^{-Mt}$$

looks like an exponential function again!
(time to first migration event back in time is exponentially distributed with mean $1/M$)

$$(4.9) \quad \Pr(\text{individual was not a migrant last generation}) = 1 - m$$

$$(4.10) \quad \Pr(\text{individual was not a migrant in the past } r \text{ generations}) = (1 - m)^r$$

setting time to be in units of $2N$ generations ($r = 2Nt$), $N \rightarrow \infty$ and $M=2Nm$ (ie $m = M/2N$), so

$$(1 - m)^r = (1 - (M/2N))^{2Nt} = e^{-Mt}$$

looks like an exponential function again!
(time to first migration event back in time is exponentially distributed with mean $1/M$)

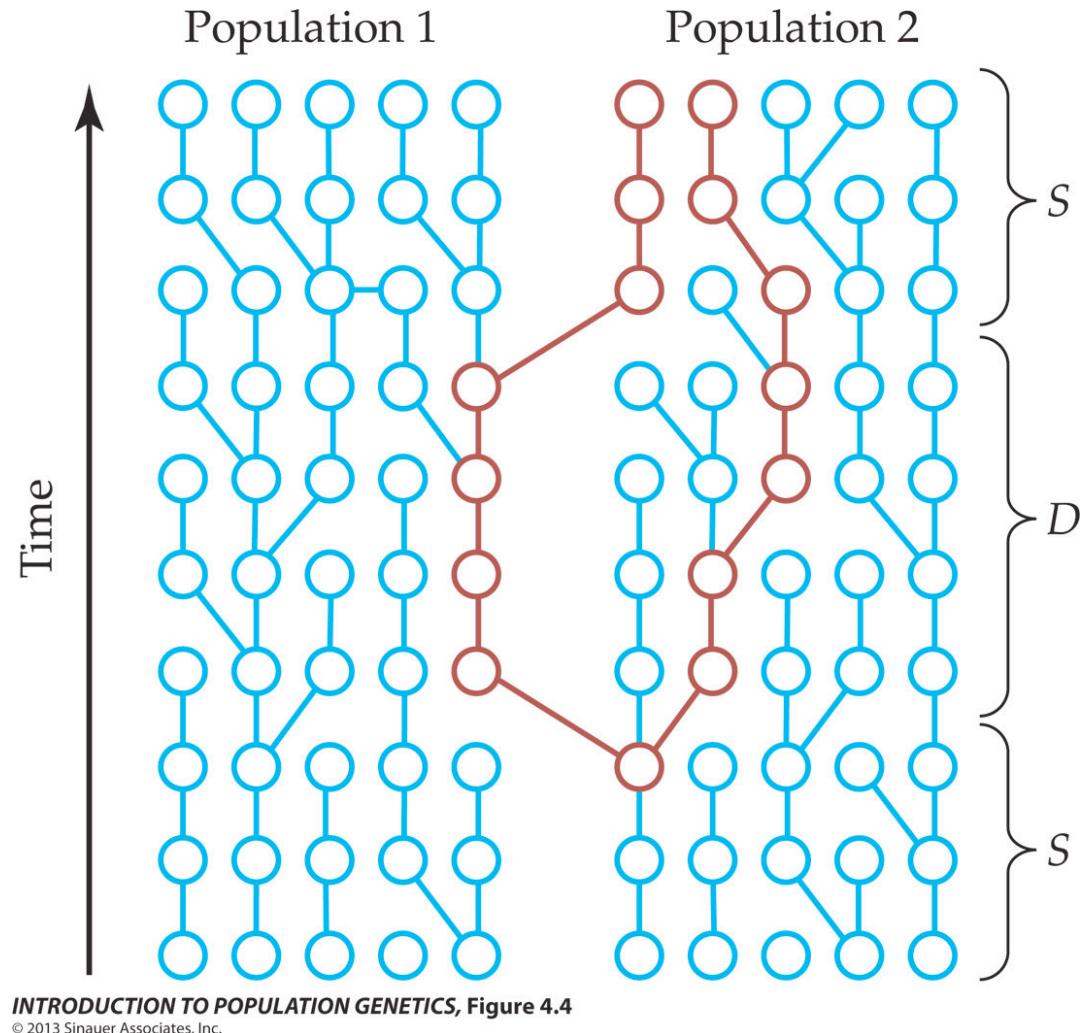
important: $M=Nm$

- M is biologically relevant and can be estimated
- m can not be estimated without knowing N

Time to coalescence given 2 lineages ($n = 2$ samples)

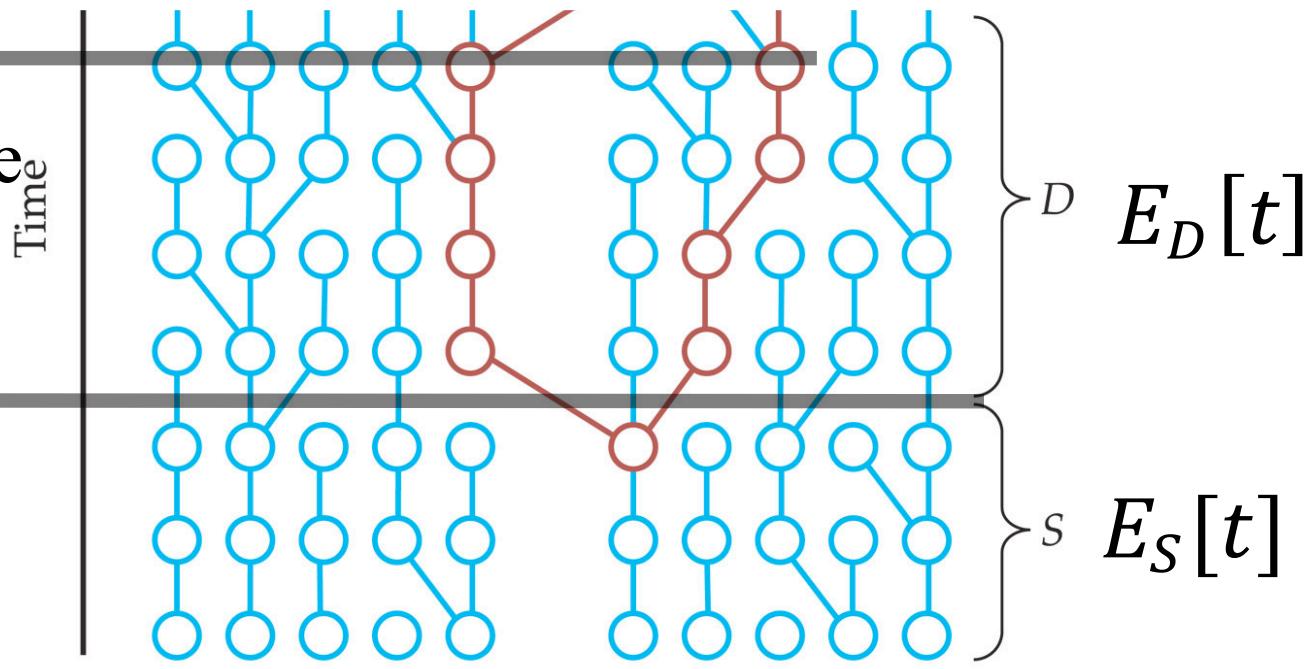
$E_D[t]$ expected time to coalescent if 2 samples in different populations

$E_S[t]$ expected time to coalescent if 2 samples in same populations



Time to coalescence given 2 lineages ($n = 2$ samples)

rate = $2M$; waiting time
for migration is $1/2M$

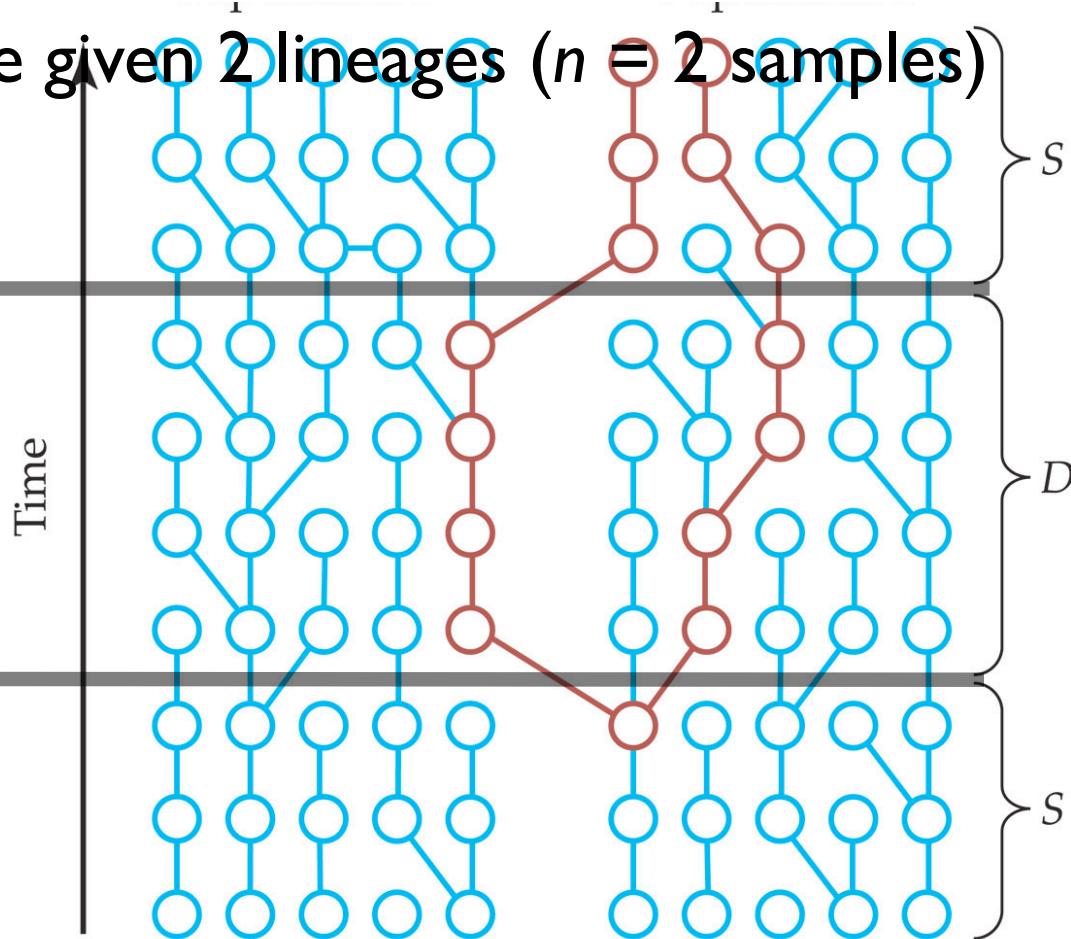


INTRODUCTION TO POPULATION GENETICS, Figure 4.4
© 2013 Sinauer Associates, Inc.

$$E_D[t] = \frac{1}{2M} + E_S[t]$$

Time to coalescence given 2 lineages ($n = 2$ samples)

$$E_S[t]$$

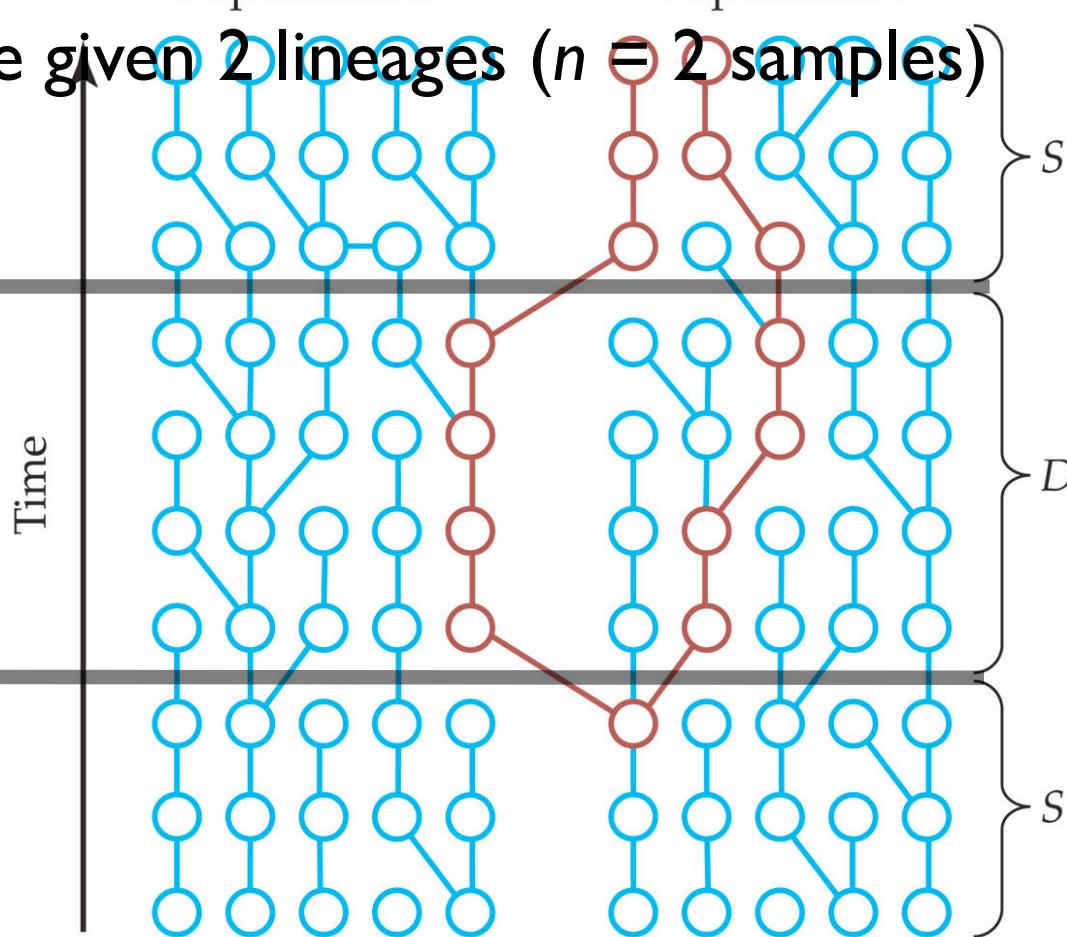


INTRODUCTION TO POPULATION GENETICS, Figure 4.4
© 2013 Sinauer Associates, Inc.

rate of either migration or coalescence = $1+2M$ (1 is rate of coalescent, and $2M$ is rate of migration) ; waiting time for either happening is $1/(1+2M)$

Time to coalescence given 2 lineages ($n = 2$ samples)

$$E_S[t]$$



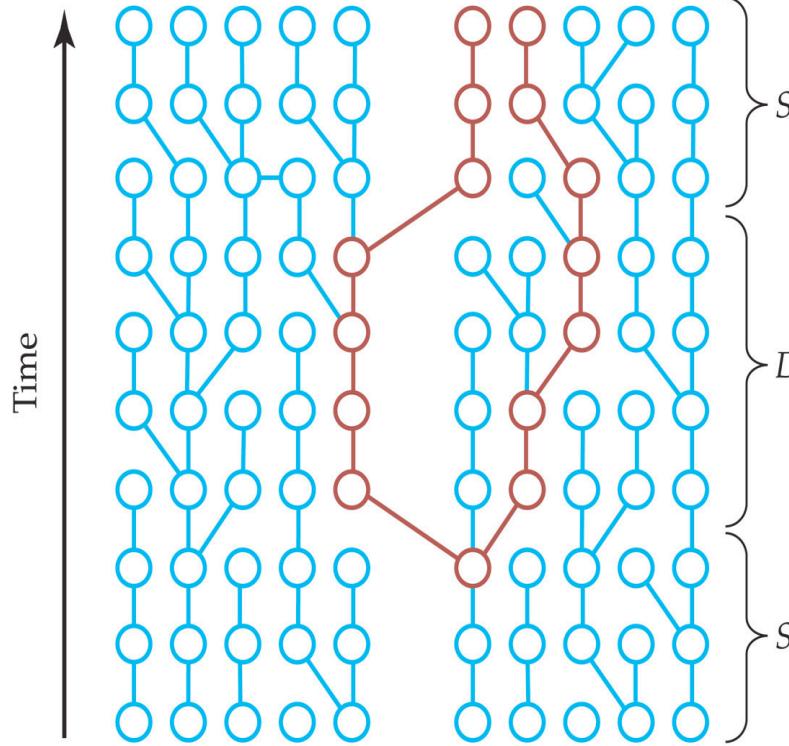
INTRODUCTION TO POPULATION GENETICS, Figure 4.4
© 2013 Sinauer Associates, Inc.

rate of either migration or coalescence = $1+2M$ (1 is rate of coalescent, and $2M$ is rate of migration ; waiting time for either happening is $1/(1+2M)$)

rate of migration and not coalescence = $2M/(1+2M)$

(ie the rate of migration, over total rate of migration plus coal)

Time to coalescence given 2 lineages ($n = 2$ samples)



rate of either migration
or coalescence

rate of migration,
but not coalescence

$$E_S[t] = \frac{1}{1 + 2M} + \frac{2M}{1 + 2M} E_D[t]$$

forcing migrant
to other
population, so
you have to
wait
 $E_D[t] = 1/2M + E_s[t]$

Time to coalescence given 2 lineages ($n = 2$ samples)

$$E_D[t] = \frac{1}{2M} + E_S[t]$$

$$E_S[t] = \frac{1}{1+2M} + \frac{2M}{1+2M} E_D[t]$$



We now have 2 equations with 2 unknowns.
Just solve this system of equations

Monty Slatkin

Time to coalescence given 2 lineages ($n = 2$ samples)

$$E_D[t] = \frac{1}{2M} + E_S[t]$$

$$E_S[t] = \frac{1}{1+2M} + \frac{2M}{1+2M} E_D[t]$$



Solving this system of equations, it is just

$$E_D[t] = \frac{1}{2M} + 2 \quad E_S[t] = 2$$

Monty Slatkin

Time to coalescence given 2 lineages ($n = 2$ samples)

$$E_D[t] = \frac{1}{2M} + E_S[t]$$

$$E_D[t] = \frac{1}{2M} + 2$$

$$E_S[t] = \frac{1}{1+2M} + \frac{2M}{1+2M} E_D[t]$$

$$E_S[t] = 2$$

Time to coalescence controlled by M

what about more than 2 pops? (many pops?)

what about more than 2 pops? (d pops?)

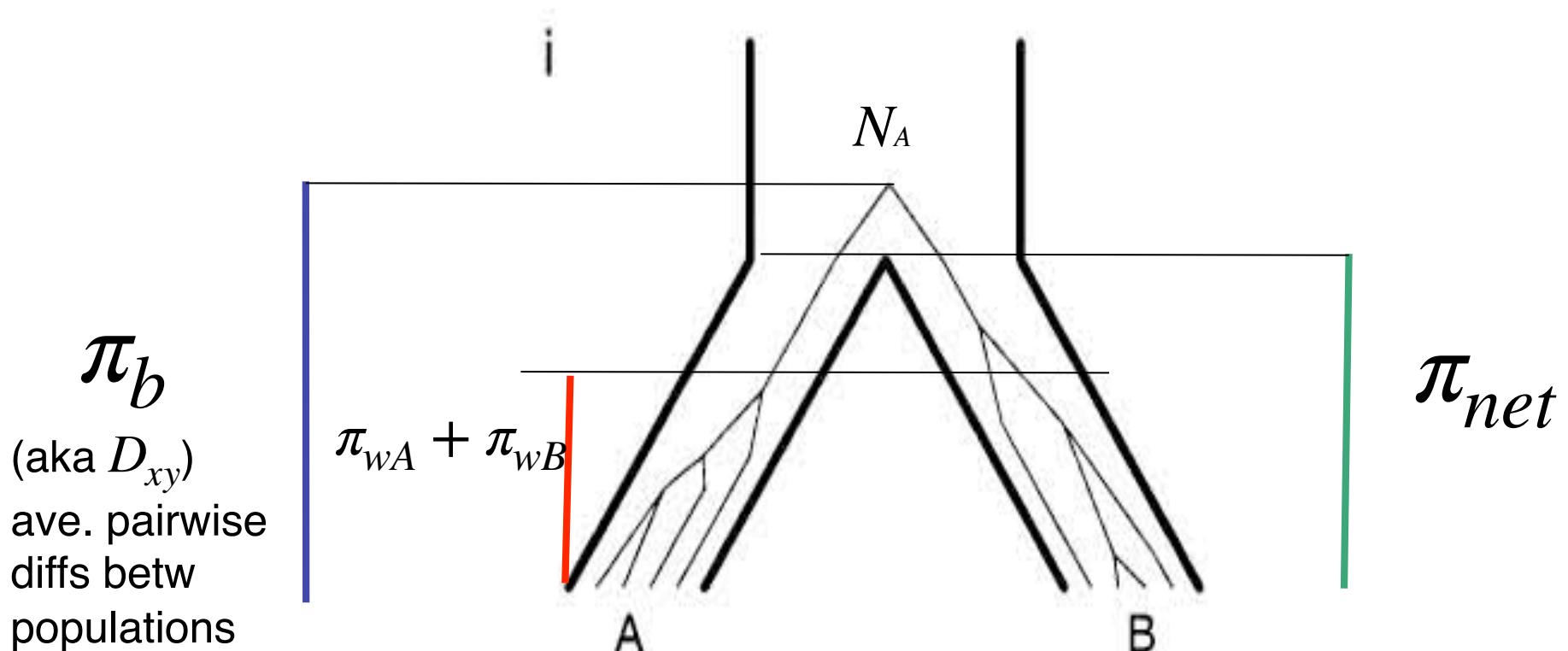


this is easily extended, a very similar calculation gives you ...

$$E_s[t] = d \quad E_D[t] = 1/2M + d$$

Monty Slatkin

Estimating Divergence time with Nei and Li's π -net



$$\pi_{net} = \pi_b - (\pi_{wA} + \pi_{wB})/2$$

