Ch 4: Population Subdivision

Population Structure

- most natural populations exist across a landscape (or seascape) that is more or less divided into areas of suitable habitat
- to the extent that populations are isolated, they will become genetically differentiated due to genetic drift, selection, and eventually mutation
- genetic differentiation among populations is relevant to conservation biology as well as fundamental questions about how adaptive evolution proceeds

Definitions

- panmixia
- population structure
- subpopulation
- gene flow
- isolation by distance
- vicariance (vicariant event)

Structure Results in "Inbreeding"

- given finite population size, autozygosity gradually increases because the members of a population share common ancestors
  - even when there is no **close** inbreeding

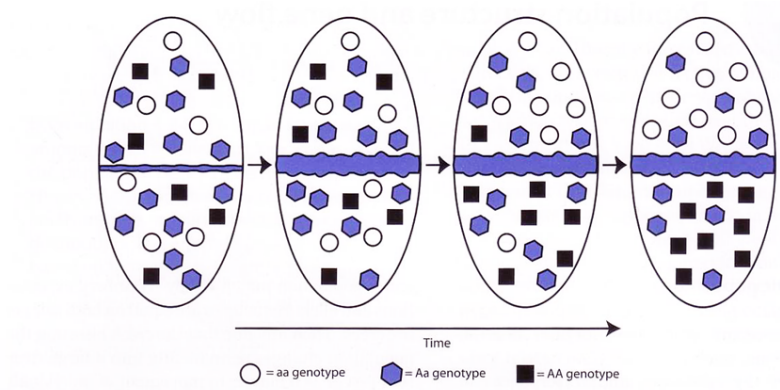


Figure 4.1 An example of population structure and allele-frequency divergence produced by limited gene flow. The total population (large ovals) is initially in panmixia and has Hardy–Weinberg expected genotype frequencies. Then the stream that runs through the population grows into a large river, restricting gene flow between the two sides of the total population. Over time allele frequencies diverge in the two subpopulations through genetic drift. In this example, you can imagine that the two subpopulations drift toward fixation for different alleles but neither reaches fixation due to an occasional individual that is able to cross the river and mate. Note that there is random mating (panmixia) within each subpopulation so that Hardy–Weinberg expected genotype frequencies are maintained within subpopulations. However, after the initial time period genotype frequencies in the total population do not meet Hardy–Weinberg expectations.

"Identical by Descent"

- what is the probability that two randomly sampled alleles are identical by descent (i.e., "replicas of a gene present in a previous generation")?
  - Wright's "fixation index" F
- at the start of the process (time 0), "declare" all alleles in the population to be unique or unrelated, $F_t = 0$ at $t = 0$
- in the next generation, the probability of two randomly sampled alleles being copies of the same allele from a single parent = $1/(2N)$, so…

"Identical by Descent"

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

= probability that alleles are copies of the same gene from the immediately preceding generation *plus* the probability that the alleles are copies of the same gene from an earlier generation

or

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

assuming $F_0 = 0$

compare to:
mean time
to fixation for
new mutant
= ~$4N$

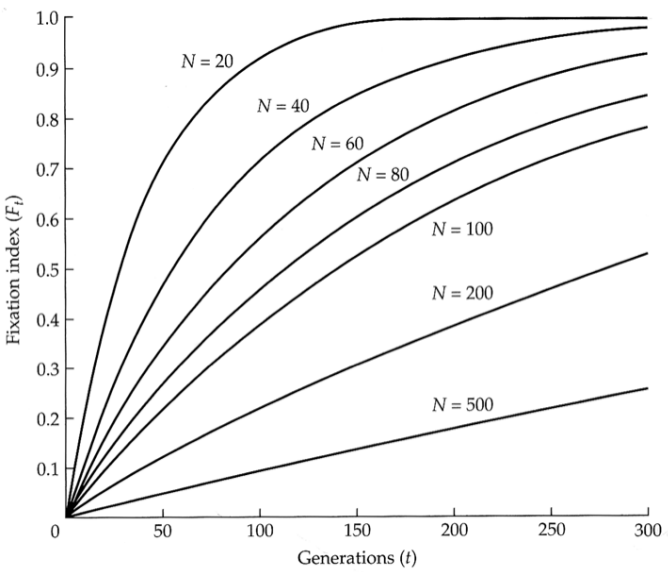
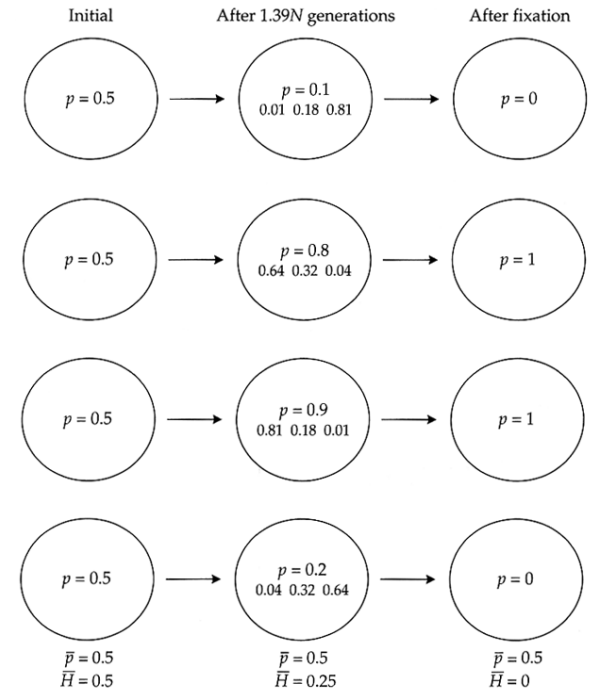


FIGURE 3.11 Increase of F_t in ideal populations as a function of time and effective population size N.

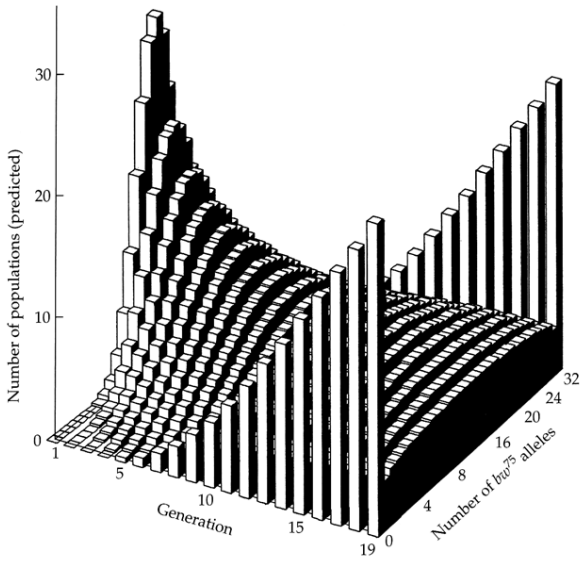
- Suppose multiple subpopulations:

Overall average allele frequency stays the same but heterozygosity declines



Predicted distributions of allele frequencies in replicate populations of $N = 16$

same process as in this figure…



Population Structure

- F_t for a single population is essentially the same thing as F_{ST}
  - a measure of genetic differentiation among populations based on the reduction in heterozygosity
- due to increasing autozygosity, structured populations have lower heterozygosity than expected if all were combined into a single random breeding population

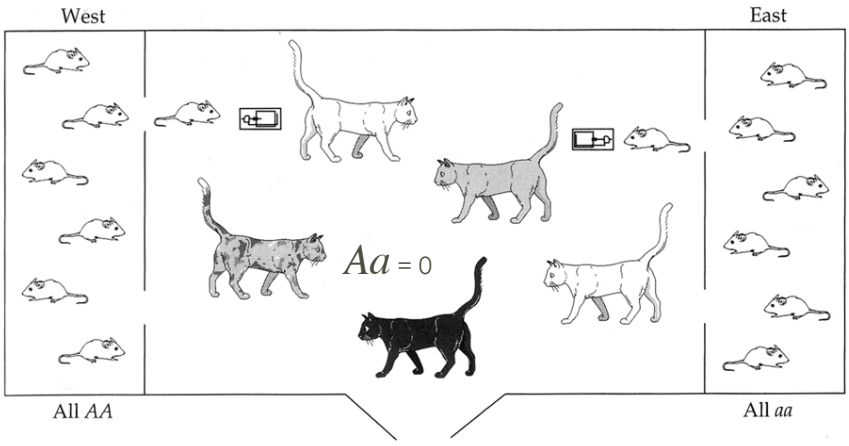


FIGURE 6.12 An extreme example of the general principle that a difference in allele frequency among subpopulations results in a deficiency of heterozygotes. The floor plan is that of a hypothetical barn. The mouse subpopulations in the east and west enclaves are completely isolated because of the cats 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.

F_{ST}

- measures the deficiency of heterozygotes in the total population relative to the expected level (assuming HWE)
- in the simplest case, one can calculate F_{ST} for a comparison of two populations…

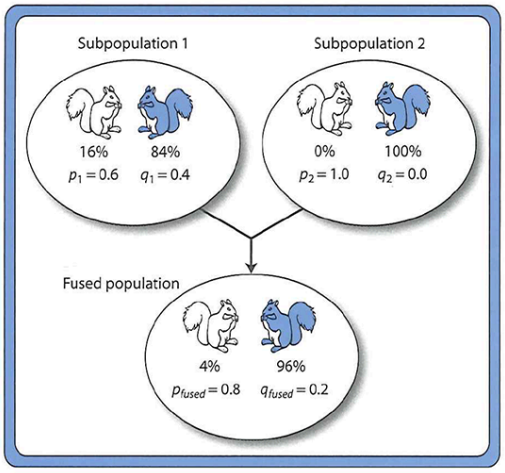
$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Two population, two allele F_{ST}

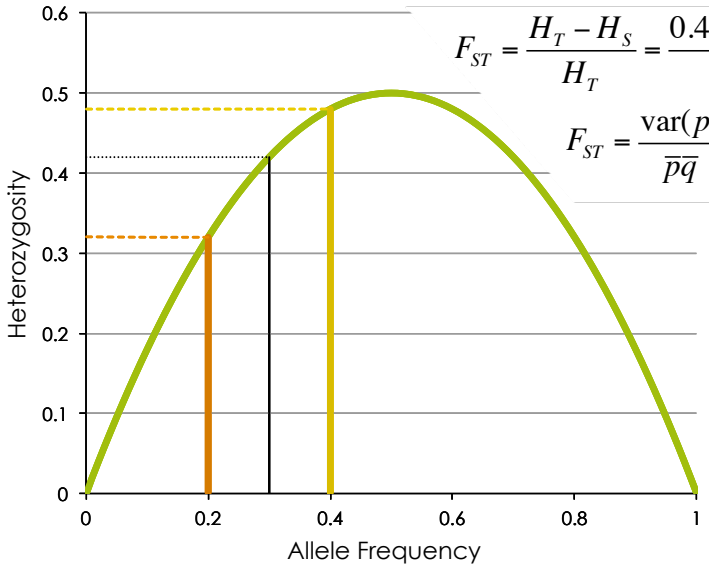
Frequency of "A"				
Population 1	Population 2	H_T	H_S	F_{ST}
0.5	0.5	0.5	0.5	0
0.4	0.6	0.5	0.48	0.04
0.3	0.7	0.5	0.42	0.16
0.2	0.8	0.5	0.32	0.36
0.1	0.9	0.5	0.18	0.64
0.0	1.0	0.5	0	1
0.3	0.35	0.43875	0.4375	0.002849
0.65	0.95	0.32	0.275	0.140625

F_{ST} - Whalund Effect

- Whalund principle - reduction in homozygosity that results from combining differentiated populations



Frequency of heterozygotes in the combined population is higher than the average of the separate populations (0.42 > 0.40)



$$F_{ST} = \frac{H_T - H_S}{H_T} = \frac{0.42 - 0.40}{0.42} = 0.0476$$

$$F_{ST} = \frac{\text{var}(p)}{\bar{p}\bar{q}} = \frac{0.01}{0.21} = 0.0476$$

F_{ST} - Whalund Effect

- Whalund principle - reduction in homozygosity due to combining differentiated populations
  - *R* = frequency of homozygous recessive genotype

$$R_{separate} - R_{fused} = \frac{q_1^2 + q_2^2}{2} - \bar{q}^2$$

$$= \frac{1}{2}(q_1 - \bar{q})^2 + \frac{1}{2}(q_2 - \bar{q})^2$$

$$= \sigma_q^2$$

F_{ST} - Whalund Effect (Nielsen & Slatkin)

$$f_A = \frac{2N_1 f_{A1} + 2N_2 f_{A2}}{2N_1 + 2N_2} \equiv f_A = \frac{f_{A1} + f_{A2}}{2}$$

$$H_S = \frac{2f_{A1}(1 - f_{A1}) + 2f_{A2}(1 - f_{A2})}{2} = f_{A1}(1 - f_{A1}) + f_{A2}(1 - f_{A2})$$

$$H_T = 2\left(\frac{f_{A1} + f_{A2}}{2}\right)\left(1 - \frac{f_{A1} + f_{A2}}{2}\right) = f_{A1}(1 - f_{A1}) + f_{A2}(1 - f_{A2}) + \frac{\delta^2}{2}$$

where $\delta = |f_{A1} - f_{A2}|$

F_{ST} over time w/ no migration

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

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t \approx 1 - e^{-\frac{1}{2N}t}$$

$$F_{ST} \approx 1 - e^{-\frac{1}{2N}t}$$

F_{ST} increases with time due to genetic drift in exactly the same way as F_t →

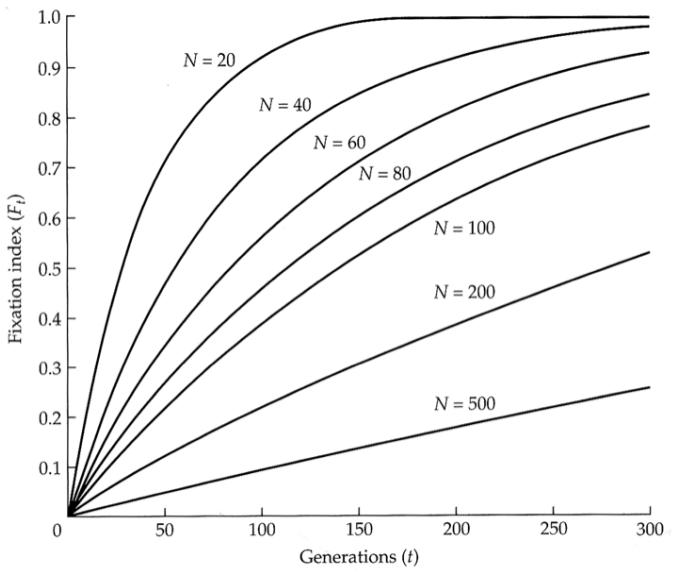


FIGURE 3.11 Increase of F_t in ideal populations as a function of time and effective population size N.

Migration

- migration between populations results in *gene flow*, which counters the effects of genetic drift (and selection) and tends to homogenize allele frequencies
- what level of migration is sufficient to counter the effects of genetic drift?
  - $Nm \sim 1$
- what level of migration is sufficient to counter the effects of selection?
  - $m > s$

The Island Model

assumptions:

- equal population sizes
- equal migration rates in all directions

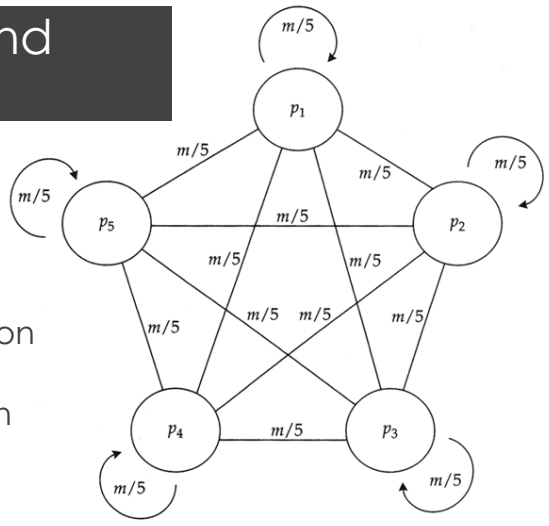


FIGURE 6.18 The island model of migration with five subpopulations. Migration is completely symmetrical. Each subpopulation contributes individuals or gametes to a pool of migrants, which then distribute themselves randomly among the subpopulations. In this model, a migrant can re-enter the same subpopulation it came from, indicated by the loops.

Equilibrium value of F_{ST}

- change in F_t with migration

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

setting $\hat{F} = F_t = F_{t-1}$

some algebra + ignoring terms in m^2 and m/N...

$$\hat{F} \approx \frac{1}{1+4Nm}$$

Equilibrium value of F_{ST}

$$\hat{F} \approx \frac{1}{1+4Nm}$$

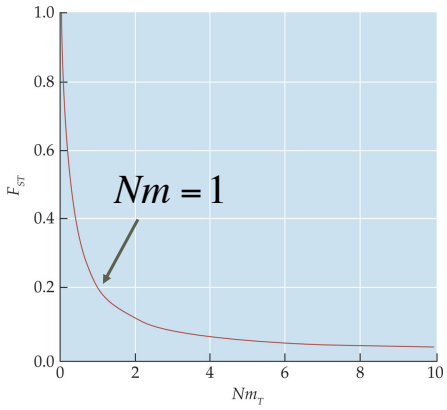


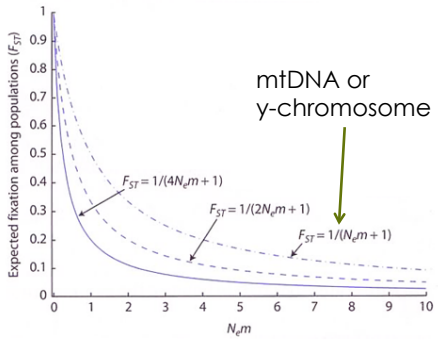
Fig. 4.5, pg. 69

Migration rate vs. Number of migrants

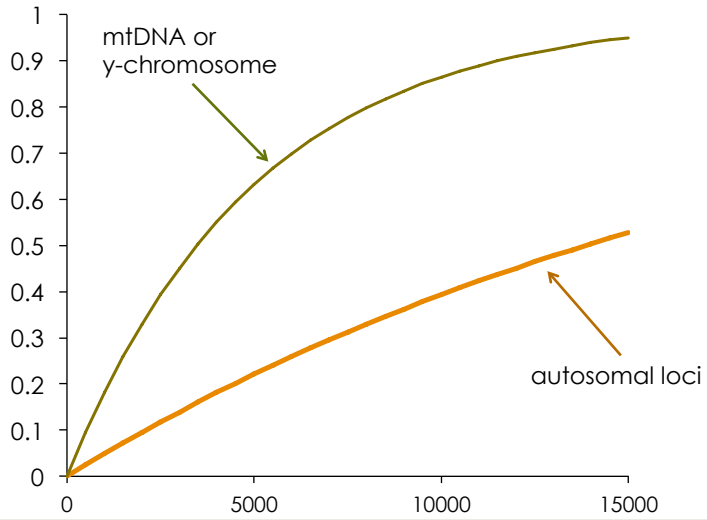
- migration rates yielding $Nm = 1$
  - N_e = 100, m = 0.01
  - N_e = 1000, m = 0.001
  - N_e = 10000, m = 0.0001
  - N_e = 100000, m = 0.00001

Equilibrium value of F_{ST}

Figure 4.15 Expected levels of fixation among subpopulations depend on the product of the effective population size (N_e) and the amount of gene flow (m) in the infinite island model of population structure. Each line represents expected F_{ST} for loci with different probabilities of autozogosity (from bottom to top $\frac{1}{2N_e}$, $\frac{1}{N_e}$, and $\frac{2}{N_e}$). Marked divergence of allele frequencies among subpopulations ($F_{ST} \geq 0.2$) are expected when $N_e m$ is below 1 for biparentally inherited nuclear loci with an autozygosity of $\frac{1}{2N_e}$. Y-chromosome or mitochondrial loci (autozygosity = $\frac{2}{N_e}$) are examples where marked divergence among populations is expected at higher levels of $N_e m$.



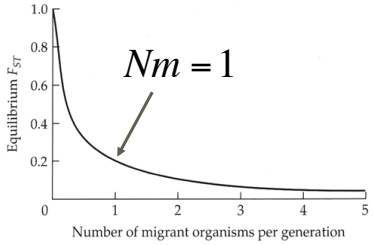
F_{ST} over time w/ no migration



Nm = 1 corresponds to F_{ST} = 0.2

- Wright (1978)
  - F_{ST} = 0.05 to 0.15 - "moderate differentiation"
  - F_{ST} = 0.15 to 0.25 - "great genetic differentiation"
  - F_{ST} > 0.25 - "very great genetic differentiation"

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.



Nm = 1 corresponds to F_{ST} = 0.2

- Wright (1978)
  - F_{ST} = 0.05 to 0.15 - "moderate differentiation"
  - F_{ST} = 0.15 to 0.25 - "great genetic differentiation"
  - F_{ST} > 0.25 - "very great genetic differentiation"
- populations of most mammalian species range from F_{ST} = 0.1 to 0.8
- humans:
  - among European groups: 0 to 0.025
  - Among Asians, Africans & Europeans: 0.05 to 0.2

F_{ST}

- theoretical maximum is 1 if two populations are fixed for different alleles
- but, there are some issues…
- fixation index developed by Wright in 1921 when we knew essentially nothing about molecular genetics
  - two alleles at a locus (with or w/o mutation between them) was the model

F_{ST} versus G_{ST}

- F_{ST} – derived by Wright as a function of the variance in allele frequencies

$$F_{ST} = \frac{\mathrm{var}(p)}{\bar{p}\bar{q}}$$

- G_{ST} – derived by Nei as a function of within and among population heterozygosities

$$G_{ST} = \frac{H_T - H_S}{H_T} = 1 - \left(\frac{H_S}{H_T} \right)$$

G_{ST} with multiple alleles

- microsatellite loci, for example, may have many alleles in all subpopulations
- F_{ST} can not exceed the average level of homozygosity (1 minus heterozygosity)

$$G_{ST} = 1 - \frac{H_S}{H_T} < 1 - H_S$$

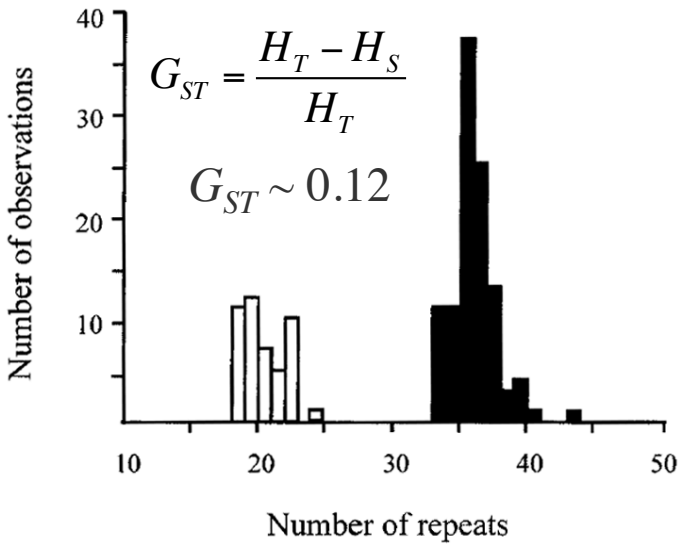


FIG. 2. Allele distribution of the Y-chromosome microsatellite L8Y. Males of the Cordon race are represented by black bars and Valais males by white bars.

Balloux *et al.* 2000 *Evolution*

Hedrick (2005) Evolution

- a standardized genetic distance measure for k populations: G'_{ST}

$$G'_{ST} = \frac{G_{ST}}{G_{ST(\mathrm{Max})}} = \frac{G_{ST}\left(k - 1 + H_S\right)}{\left(k-1\right)\left(1 - H_S\right)}$$

- where:

$$G_{ST(\mathrm{Max})} = \frac{H_{T(\mathrm{Max})} - H_S}{H_{T(\mathrm{Max})}} \quad \text{and} \quad H_{T(\mathrm{Max})} = 1 - \frac{1}{k^2}\sum_i \sum_j p_{ij}^2$$

TABLE 1. Examples illustrating the effect of heterozygosity on measures of genetic differentiation. (a), (b), and (c) have $H_S = 0.25$ and $G_{ST(max)} = 0.6$ whereas (d), (e), and (f) have $H_S = 0.58$ and $G_{ST(max)} = 0.266$.

	(a) Subpopulation		(b) Subpopulation		(c) Subpopulation	
Allele	1	2	1	2	1	2
1	0.1	—	0.1	—	0.9	—
2	0.9	—	0.9	0.8	0.1	0.2
3	—	0.8	—	0.2	—	0.8
4	—	0.2	—	—	—	—
H_S	0.25		0.25		0.25	
$G_{ST(max)}$	0.6		0.6		0.6	
G_{ST}	0.6		0.057		0.593	
G'_{ST}	1.0		0.095		0.988	

	(d) Subpopulation		(e) Subpopulation		(f) Subpopulation	
Allele	1	2	1	2	1	2
1	0.5	—	0.1	—	0.6	—
2	0.5	—	0.6	0.5	0.3	—
3	—	0.3	0.3	0.3	0.1	0.2
4	—	0.3	—	0.2	—	0.3
5	—	0.4	—	—	—	0.5
H_S	0.58		0.58		0.58	
$G_{ST(max)}$	0.266		0.266		0.266	
G_{ST}	0.266		0.025		0.256	
G'_{ST}	1.0		0.094		0.964	

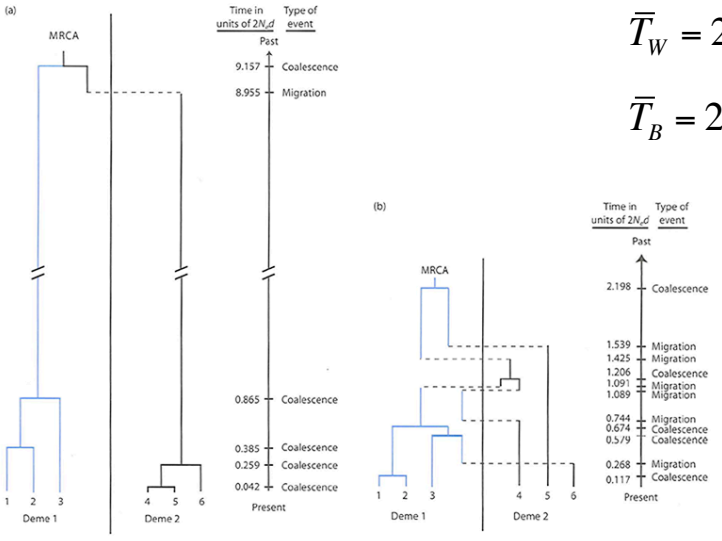
	Subpopulation			Subpopulation	
Allele	**1**	**2**		**1**	**2**
1	0.1	—		0.1	—
2	0.2	—		0.2	—
3	0.2	—		0.2	0.1
4	0.2	—		0.2	0.2
5	0.2	—		0.2	0.2
6	0.1	—		0.1	0.2
7	—	0.1		—	0.2
8	—	0.2		—	0.1
9	—	0.2		—	—
10	—	0.2		—	—
11	—	0.2		—	—
12	—	0.1		—	—
H_S	0.820		H_S	0.820	
H_T	0.910		H_T	0.850	
F_{ST} (G_{ST})	**0.099**		F_{ST} (G_{ST})	**0.035**	
$H_{T(max)}$	0.910		$H_{T(max)}$	0.910	
$G_{ST(max)}$	0.099		$G_{ST(max)}$	0.099	
G'_{ST}	**1**		G'_{ST}	**0.357**	

Coalescent-based Measures

- Slatkin (1995) *Genetics*

$$F_{ST} = \frac{\bar{T} - \bar{T}_W}{\bar{T}}$$

- where $\bar{T}$ and $\bar{T}_W$ are the mean coalescence times for all alleles and alleles within subpopulations



$$\bar{T}_W = 2N_e d$$

$$\bar{T}_B = 2N_e d + \frac{d-1}{2m}$$

Figure 4.19 Genealogies for six lineages initially divided evenly between two demes when the migration rate is low (a) and when the migration rate is high (b). When migration is unlikely, coalescent events within demes tend to result in a single lineage within all demes before any migration events take place. There is then a long wait until a migration event places both demes in one deme where they can coalesce. When migration is likely, lineages regularly move between the demes, and lineages originally in the same deme are as likely to coalesce as lineages initially in different demes. These two genealogies are examples and substantial variation in coalescence times is expected. In (a) $M = 4N_e m = 0.2$ and in (b) $M = 4N_e m = 2.0$. The two genealogies are not drawn to the same scale. MRCA, most recent common ancestor.

R_{ST} for microsatellites

- under a stepwise mutation model for microsatellites, the difference in repeat number is correlated with time to coalescence

$$R_{ST} = \frac{\bar{S} - S_W}{\bar{S}}$$

- where $\bar{S}$ and $\bar{S}_W$ are the average squared difference in repeat number for all alleles and alleles within subpopulations
- violations of the stepwise mutation model are a potential problem

Φ_{ST} for DNA sequences

- the number of pairwise differences between two sequences provides an estimate of time to coalescence
- method of Excoffier *et al*. (1992) takes into account the number of differences between haplotypes
- *Arelquin* (software for AMOVA analyses) calculates both F_{ST} and Φ_{ST} for DNA sequence data
  - important to specify which one is calculated