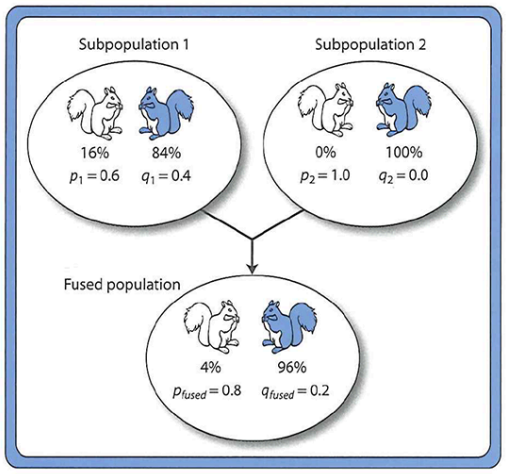
F_{ST} - Whalund Effect

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



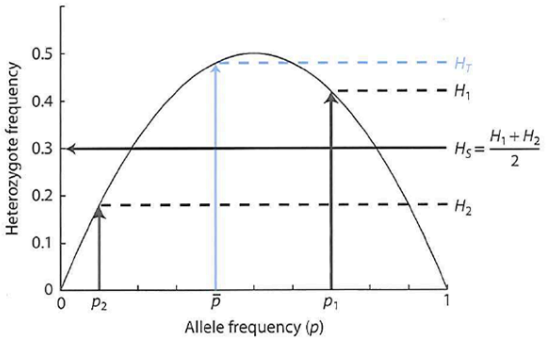
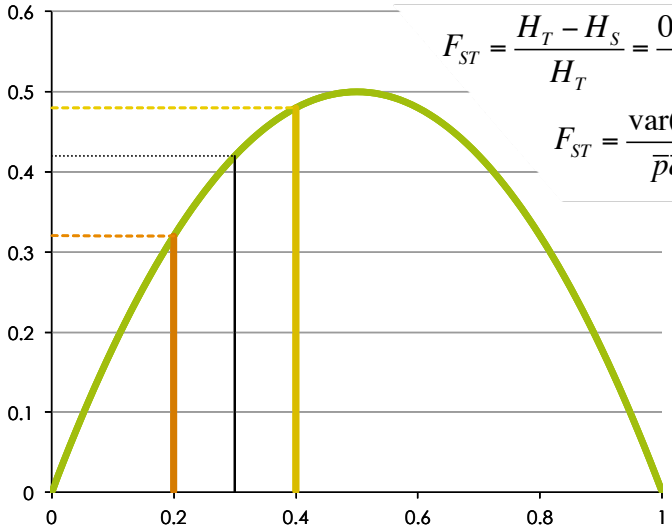


Figure 4.10 A graphical demonstration of the Wahlund effect for a diallelic locus in two demes. If there is random mating within subpopulations (H_1 and H_2) and in the total population (H_T), the heterozygosity of each falls on the parabola of Hardy–Weinberg expected frequency. The average heterozygosity of subpopulations (H_S) is at the mid-point between the deme heterozygosities. Therefore, H_S can never be greater than H_T based on the average allele frequency (the mid-point between the deme allele frequencies p_1 and p_2). Greater variance in allele frequencies of the demes is the same as a wider spread of deme allele frequencies in the two-deme case.

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



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

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

pg. 129

Table 4.7	Initial subpopulations	Fused population
Variance in q	$\frac{(0.4-0.2)^2+(0.0-0.2)^2}{2} = 0.04$	0

$$\text{Var}(q) = \frac{(0.4-0.2)^2+(0.0-0.2)^2}{(2-1)} = 0.08 \quad (4.36)$$

whereas var(q) is zero after fusion because there is no longer any subdivision for allele frequencies. Take note of the fact that the initial variance in the allele frequencies (0.08) is exactly twice the difference between the average frequency of albinos before fusion (equation 4.32) and the expected frequency of albinos in the fused population (equation 4.34)! With fusion of the subdivided populations, each homozygote has decreased by 4% and the heterozygote has increased by exactly the same total amount or 8%.

	pop1-q	pop2-q	pop3-q	avg	FUSED	diff(avg-fused)	var(q)	
q	0.4	0		0.20	0.20		0.04	n
q^2	0.16	0		0.08	0.04	0.04	0.08	(n-1)

	pop1-q	pop2-q	pop3-q	avg	FUSED	diff(avg-fused)	var(q)	
q	0.4	0	0.25	0.2167	0.2167		0.0272	n
q^2	0.16	0	0.0625	0.0742	0.0469	0.0272	0.0408	(n-1)

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}

- 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)

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

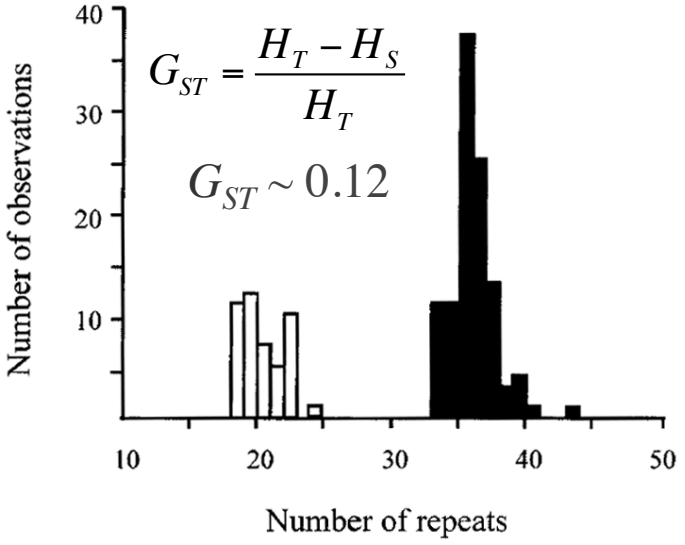


FIG. 2. Allele distribution of the Y-chromosome microsatellite L8Y. Males of the Cordon race are represented by black bars and Valais males by white bars.

Balloux *et al.* 2000 *Evolution*

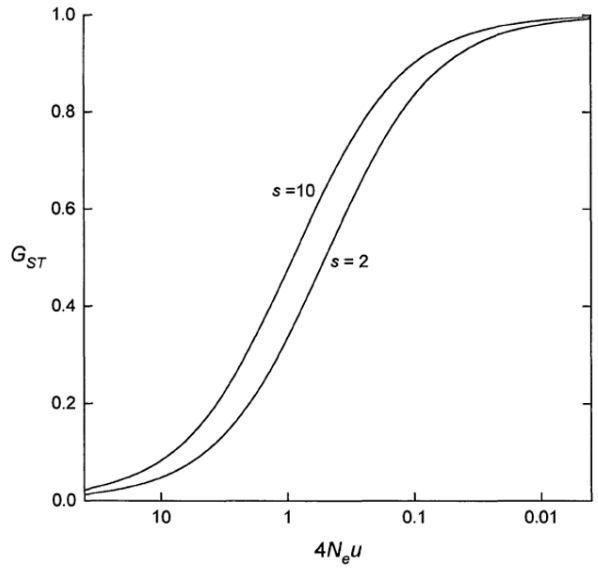


FIG. 1. The asymptotic level of G_{ST} between completely isolated populations as a function of the mutation rate for two and 10 populations.

Hedrick 1999
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*

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

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

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

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

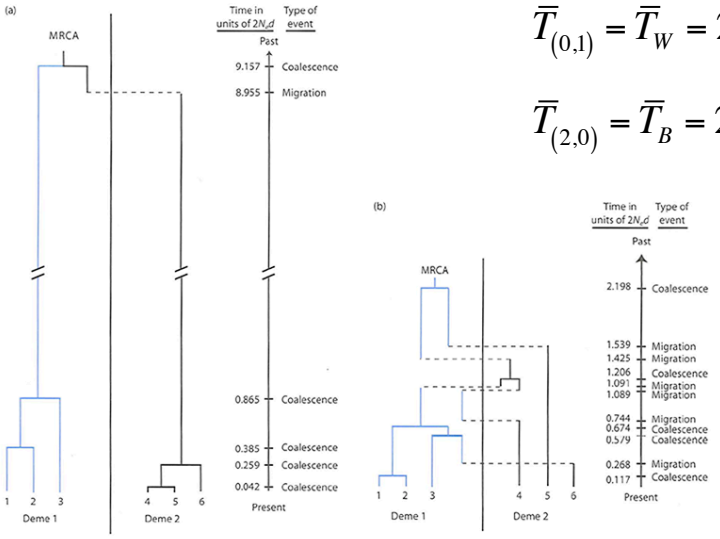


Figure 4.19 Genealogies for six lineages initially divided evenly between two demes when the migration rate is low (a) and when the migration rate is high (b). When migration is unlikely, coalescent events within demes tend to result in a single lineage within all demes before any migration events take 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