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
“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”

\[ F_i = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{i-1} \]

= 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

\[ F_i = 1 - \left(1 - \frac{1}{2N}\right)^i \] assuming \( F_0 = 0 \)

compare to:
mean time to fixation for new mutant
= \( \sim 4N \)

**FIGURE 3.11** Increase of \( F_i \) 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

- $F$, 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 $F_{ST}$

<table>
<thead>
<tr>
<th>Frequency of &quot;A&quot;</th>
<th>Population 1</th>
<th>Population 2</th>
<th>$H_T$</th>
<th>$H_S$</th>
<th>$F_{ST}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.48</td>
<td>0.04</td>
</tr>
<tr>
<td>0.3</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.42</td>
<td>0.16</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8</td>
<td>0.5</td>
<td>0.5</td>
<td>0.32</td>
<td>0.36</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.18</td>
<td>0.64</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.3</td>
<td>0.35</td>
<td>0.43875</td>
<td>0.4375</td>
<td>0.4375</td>
<td>0.002849</td>
</tr>
<tr>
<td>0.65</td>
<td>0.95</td>
<td>0.32</td>
<td>0.275</td>
<td>0.140625</td>
<td>0.140625</td>
</tr>
</tbody>
</table>
**$F_{ST}$ - Whalund Effect**

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

![Diagram showing allele frequencies in subpopulations and the fused population.](image)

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

$$\text{Frequency of heterozygotes} = H_T - H_S$$

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

$$F_{ST} = \frac{\text{var}(p)}{pq} = \frac{0.01}{0.21} = 0.0476$$
**F<sub>ST</sub> - Whalund Effect**

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

\[
R_{\text{separate}} - R_{\text{fused}} = \frac{q_1^2 + q_2^2}{2} - \bar{q}^2 = \frac{1}{2}(q_1 - \bar{q})^2 + \frac{1}{2}(q_2 - \bar{q})^2 = \sigma_q^2
\]

**F<sub>ST</sub> - Whalund Effect** (Nielsen & Slatkin)

\[
f_A = \frac{2N_1 f_{A1} + 2N_2 f_{A2}}{2N_1 + 2N_2} = \frac{f_{A1} + f_{A2}}{2}
\]

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

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

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 autosomy (from bottom to top $\frac{1}{2N_e}, \frac{1}{N_e}, \text{ and } \frac{2}{N_e}$). Marked divergence of allele frequencies among subpopulations ($F_{ST} \geq 0.2$) are expected when $Nm$ is below 1 for biparentally inherited nuclear loci with an autoxosity of $\frac{1}{2N_e}$. Y-chromosome or mitochondrial loci (autoxosity $= \frac{2}{N_e}$) are examples where marked divergence among populations is expected at higher levels of $N_e m$. 

mlDNA or y-chromosome
$F_{ST}$ 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 6.21, giving $F_{ST}$ as a function of $Nm$. In the island model, $Nm$ is the number of migrant organisms that come into each subpopulation in each generation.
\[ Nm = 1 \text{ corresponds to } F_{ST} = 0.2 \]

- 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
  \[ F_{ST} = \frac{\text{var}(p)}{pq} \]

- $G_{ST}$ – 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)
  \[ G_{ST} = 1 - \frac{H_S}{H_T} < 1 - H_S \]
Hedrick (2005) Evolution

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

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

- where:

$$G_{ST(\text{Max})} = \frac{H_{T(\text{Max})} - H_S}{H_{T(\text{Max})}}$$

and

$$H_{T(\text{Max})} = 1 - \frac{1}{k^2} \sum_i \sum_j p_{ij}^2$$

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
### 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 $G_{ST(max)} = 0.58$.

<table>
<thead>
<tr>
<th>Allele</th>
<th>(a) Subpopulation</th>
<th>(b) Subpopulation</th>
<th>(c) Subpopulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>—</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>—</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>0.2</td>
<td>—</td>
</tr>
<tr>
<td>$H_S$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>$G_{ST(max)}$</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>$G_{ST}$</td>
<td>0.095</td>
<td>0.983</td>
<td>0.983</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Allele</th>
<th>(d) Subpopulation</th>
<th>(e) Subpopulation</th>
<th>(f) Subpopulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>—</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>0.2</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>0.4</td>
<td>—</td>
</tr>
<tr>
<td>$H_S$</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>$G_{ST(max)}$</td>
<td>0.266</td>
<td>0.266</td>
<td>0.266</td>
</tr>
<tr>
<td>$G_{ST}$</td>
<td>0.094</td>
<td>0.964</td>
<td>0.964</td>
</tr>
</tbody>
</table>

### Subpopulation

<table>
<thead>
<tr>
<th>Allele</th>
<th>Subpopulation</th>
<th>Subpopulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>0.1</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>0.2</td>
</tr>
<tr>
<td>9</td>
<td>—</td>
<td>0.2</td>
</tr>
<tr>
<td>10</td>
<td>—</td>
<td>0.2</td>
</tr>
<tr>
<td>11</td>
<td>—</td>
<td>0.2</td>
</tr>
<tr>
<td>12</td>
<td>—</td>
<td>0.1</td>
</tr>
</tbody>
</table>

| $H_S$ | 0.820         | 0.820         |
| $H_T$ | 0.910         | 0.850         |
| $F_{ST}$ ($G_{ST}$) | **0.099** | **0.035** |
| $H_{T(max)}$ | 0.910         | 0.910         |
| $G_{ST(max)}$ | 0.099         | 0.099         |
| $G_{ST}$ | **1**         | **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 each deme 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 lineages into one deme where they can coalesce. When migration is likely, lineages regularly move between the demes, and lineages originating in the same deme are less likely to coalesce as lineages initially in different demes. These two genealogies are examples of substantial variation in coalescence times expected. In (a) $M = 0.92$, $d = 0.2$, and in (b) $M = 4.92$, $d = 2.0$. The two genealogies are not drawn to the same scale. MRCA, most recent common ancestor.
**R\textsubscript{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 \( S_W \) 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

**\Phi_{ST} for DNA sequences**

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