Species, Subspecies and ESU's in Conservation Policy - see O'Brien & Mayr 1991 Crandall et al 2000.

Several points emerge from understanding of speciation and classification:

1. Under the biological species concept (BSC), the definition of a species is based on gene flow within a species and reproductive isolation among species.
2. But gene flow and (especially) reproductive isolation are generally not tested in classification.
3. Instead, cladistic and evolutionary classification rely on patterns of similarity to build phylogenetic trees based on genetic distance, and classify based on these trees.
4. Molecular genetic data on selectively neutral genetic variation is now easy to obtain. Consequently, most phylogenetic trees are now based largely or wholly on such data.

How to assess gene flow and reproductive isolation?

1. Fixation indices or F-statistics (Sewall Wright 1922), particularly $\boldsymbol{F_{ST}}$

F statistics describe how genetic variation (measured by heterozygosity, H) is fixed at each level of organization in a structured population (or subspecies, or species — the same logic can be applied at various levels)

F stats are based on H. H is a good measure of genetic variability
- increases with more alleles
- increases with more even distribution of alleles.

Classical approach is to use three levels:

Individual (I)
Subpopulation (S)
Total population (T)

a) Determine genotypes for a sample of individuals in each subpopulation.

b) From the genotypes, calculate *observed* average individual heterozygosity within a subpopulation, averaged across all subpopulations:

$$H_I = \frac{1}{k}\sum_{j=1}^{k}\frac{\#\,heterozygotes}{\#\,individuals}$$

k = number of subpopulations

c) Also calculate allele frequencies from the genotypes. Allele frequencies can be used to determine the *expected* heterozygosity for the average subpopulation and for the total population:

$$\bar{H}_s = \frac{2}{k}\sum_{j=1}^{k}(p_j - p_j^2)$$

- this is ***expected*** heterozygosity within subpops if individuals mate at random ***within*** subpops.

$H_T = 2(\bar{p} - \bar{p}^2)$ (note $\bar{p}$ is average across ALL inds)

- this is ***expected*** heterozygosity if individuals mate at random ***across*** subpopulations.

d) How much genetic variation is fixed at levels of individuals and subpopulations, relative to total variation across all subpops?

To answer this, use observed and expected H values to calculate F statistics.

$$F_{IS} = \frac{\bar{H}_s - H_I}{\bar{H}_s}$$

$F_{IS} > 0$: individuals are less variable than expected for mean level of allelic variation observed in subpopulations (evidence for positive ***assortative mating***, especially ***inbreeding***).

$$F_{IT} = \frac{H_T - H_I}{\bar{H}_T}$$

$F_{IT} > 0$: individuals are less variable than expected for level of allelic variation seen in total population

- less than full mixture of individuals across all subpops
- due to combination of all processes *within* subpop and *among* subpops
- within subpops = assortative mating & inbreeding
- among subpops = restricted gene flow

$$F_{ST} = \frac{H_T - \bar{H}_S}{H_T}$$

measures genetic differentiation among subpopulations

F_{ST} ranges from 0 to 1.
$F_{ST} = 0$: all subpopulations just as variable as total population. No differentiation.
$F_{ST} = 1$: all subpopulations fixed for alleles unique to that subpopulation.

$F_{ST} \rightarrow 1$: stronger evidence that populations are ESUs.