

Concept	Definition	Why is it important?
Allele number	Number of unique DNA sequences at a locus	Used to calculate Polymorphism and Heterozygosity
Allele frequency	Relative frequency of a specific sequence at any given locus in a population Calculated using HWE = $p^2+2pq+q^2$	<ul style="list-style-type: none"> - Measures genetic diversity - For a population or species Can provide insights into the Likelihood of genetic drift Evolutionary potential Inbreeding - Can be affected by selection, population size, mutation, migration, genetic drift, demographic stochasticity
Polymorphism	Multiple alleles at any given locus Percent loci with more than one allele	Provides information on: Potential for genetic drift Genetic diversity/variation “Health” of an individual/population/species
Heterozygosity	Percent of individuals with more than one allele at a given locus H = percent heterozygotic individuals at a locus \bar{H} = percent individuals that are heterozygotic across loci	Gives insight into the “health” of a population or species using F stats Provides information about mating system Could be used to calculate N_e N_e the size of an ideal population given the observed inbreeding and genetic drift in a real population However, not always accurate...use with caution

Concept	Calculation	Additional Points
Hardy-Weinberg Equilibrium	$1 = p^2 + 2pq + q^2$ $\chi^2 = \sum \frac{(obs\ freq - exp\ freq)^2}{exp\ freq}$	Describes the fixed equilibrium of an allele frequency in a simple population free of evolutionary forces. Assumptions: Large population size No migration No mutation No selection Random mating Diploid organism Sexual reproduction Equal allele frequencies in the sexes Non-overlapping generations
Linkage Disequilibrium	$D = x_{11} - p_1q_1$	Non-random association of alleles at different loci. The difference between the observed and expected frequencies. If loci are in HWE Calculate using allele frequencies If loci are not in HWE Calculate using genotype frequencies

F-statistics are derived from the inbreeding coefficient, f . Developed for DIPLOID organisms	-W/in an individual, Pr(2 randomly chosen alleles are Identical By Descent) - proportionate reduction in heterozygosity in inbred pop relative to reference pop $f = (H_r - H_f) / H_r$
Sewell Wright (1951) proposed a family of statistics based on f to identify deviations from Heterozygosity	F_{IS} - division within subpopulations F_{ST} - division due to subpopulations F_{IT} - deviation in the total population
F_{ST} is the most commonly used of the 3, although each has its place	Used to elucidate population structure and level of gene flow High value= more subdivision, less gene flow Low value= less subdivision, more gene flow
F_{ST} is affected by sampling scheme, highly polymorphic markers, number of populations sampled, gene flow vs. N_e , high mutation rates, etc.	Several variations: ⊖ Weir & Cockerham 1984 G_{ST} Nei (several times), Hedrick 2005 R_{ST} Slatkin 1995

Papers worth reading:

Weir and Cockerham 1984, Estimating F-statistics for the evaluation of population structure. Evolution 38:66

Hedrick 2005. A Standardized genetic differentiation measure. Evolution 59:8

Slatkin 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139 (457-462)

- I Genetic Distance- Essentially a measure of dissimilarity between Populations, Individuals, Families, Species
 - A Calculated using Genotypes, Allele frequencies, Allele sizes
 - B Used to measure population structure, gene flow, reconstruct phylogenies
 - C Comes in many flavors, depending on models of evolution, demographic history, types of markers
 - i D_{SA} - Proportion of Shared Alleles (Bowcock 1994)
 - ii D_S - Nei's Distance (Nei 1987)
 - iii \square - Angular Distance (Cavalli-Sforza and Edwards 1967)
 - iv D_{CH} - Chord Distance
 - v S_R - Sum of Squares of difference in Allele Sizes (Slatkin 1995)
 - vi $(\square\square)^2$ - Delta mu squared (Goldstein *et al.* 1995)
- etc, etc, etc...*
- II Mantel's Test- Test for correlation between matrices (Mantel 1967)

- A Used when autocorrelation among variables precludes conventional parametric approaches
- B Matrices consist of various distances or dissimilarities between all pairwise combinations of samples
- C Significance assessed by permutations
- D The beauty lies in the versatility
- i Simple Mantel's Test between distance matrices
 - ii Simple Mantel's Test between distance matrix and predictor matrix
 - iii Simple Mantel's Test between observed matrix and theoretical matrix
 - iv Partial Mantel's Test between three matrices (Smouse *et al.* 1986)
etc, etc, etc...
- E See also Legendre and Fortin 1989, Borcard *et al.* 1992, Legendre and Legendre 1998, Manly 1986, 1991, 1997 amongst oh so many others...

Statistical Methods – The Basics

Frequentist

- Probability is a long-term statement
- Asks: do I reject/accept H_a ?
- 2-hypotheses at a time
- P-value may be confounded by sample size
- No prior
- Less computationally demanding

Max Likelihood

- $P(X|H)$ – prob of data given the hypothesis
- Uses many point estimates to try to find max
- Choice of model/hypothesis important
- No prior
- Computationally demanding

Bayesian

- $P(H|X)$ – posterior prob of hypothesis given the data
- Includes priors
- Posterior prob is statement of confidence
- Influence of prior confounded by sample size
- Choice of prior subjective
- Computationally demanding