

Papers for 11 September

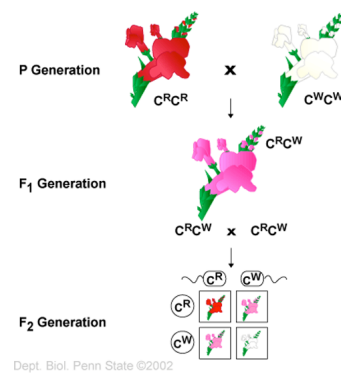
- ❖ Kreitman M (1983) Nucleotide polymorphism at the alcohol-dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412-417.
- ❖ Hishimoto *et al.* (2010) Alcohol and aldehyde dehydrogenase polymorphisms and risk for suicide: a preliminary observation in the Japanese male population. *Genes, Brain and Behavior* **9**, 498-502.
- ❖ assignment: test genotype frequencies in this paper for departures from HW-equilibrium, test each sample group separately

Chapter 1 –Allele Frequencies, Genotype Frequencies & HW Equilibrium

Alleles & Allele Frequencies
Genotype Frequencies
Hardy-Weinberg Equilibrium
Deviations from HW-equilibrium

Mendel's "Laws"

- ❖ Independent segregation
 - ❖ two members of a gene pair (alleles) segregate separately into gametes so that half of the gametes carry one allele and the other half carry the other allele
- ❖ Independent assortment
 - ❖ during gamete formation, the segregation of alleles of one gene is independent of the segregation of alleles of another gene



"Population"

- ❖ "a group of organisms of the same species living within a ... restricted geographical area so that any member can potentially mate with any other member of the same species"
 - ❖ Hartl & Clark, 2006
- ❖ "a population is a group of plants, animals, or other organisms, all of the same species, that live together and reproduce"
 - ❖ N.J. Gotelli, 1998
- ❖ "local population" = "deme" = "subpopulation"
 - ❖ fundamental unit of population genetics

Hardy-Weinberg Assumptions

- ❖ theoretical population genetics relies on a set of simplifying assumptions about the structure of populations
 - ❖ random mating* (random union of gametes)
 - ❖ non-overlapping generations*
 - ❖ migration (gene flow), mutation, and natural selection have negligible effects
 - ❖ **“large”** population size (= no genetic drift, no effect of sampling on allele frequencies)
 - ❖ sexual reproduction
 - ❖ two alleles at a locus (diploid)
 - ❖ allele frequencies equal in two sexes

Hardy-Weinberg Equilibrium (1908)

- ❖ relates allele and genotype frequencies assuming random mating
- ❖ Hardy considered the principle obvious...
- ❖ but presented it to refute an argument that 3:1 phenotypic ratios should be common in nature if inheritance was really Mendelian
 - ❖ phenotypic ratios depend on allele frequencies
- ❖ also shows that genetic variation is maintained in the absence of selection, drift, mutation
 - ❖ contrast to earlier notions of “blending inheritance”

Allele and Genotype Frequencies

❖ Allele frequencies: $f_A = \frac{N_A}{2N}$ and $f_a = \frac{N_a}{2N}$
 $f_A + f_a = 1$

❖ “evolution” = change in population allele frequencies over time

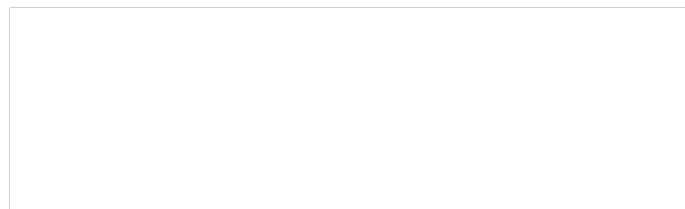
❖ Genotype frequencies:

$$f_{AA} = \frac{N_{AA}}{N} \quad f_{Aa} = \frac{N_{Aa}}{N} \quad f_{aa} = \frac{N_{aa}}{N}$$

Hardy-Weinberg Equilibrium (1908)

❖ For a population in H-W equilibrium, observed **genotype** frequencies can be predicted from the observed **allele** frequencies

$$f_A^2 + 2f_A f_a + f_a^2 = (f_A + f_a)^2 = 1$$



- ❖ Equations above effectively separate life history into two stages:
 - ❖ gametes combining to form zygotes
 - ❖ zygotes maturing into adults which produce more gametes
- ❖ more complex models including selection, drift, mutation, etc., still apply the HW principle in the formation of zygotes each generation

H-W Inferences:

- ❖ random mating of genotypes yields the same result as random union of gametes (See Box 1.2, pg. 11 - do the math!)
- ❖ a population out of HW equilibrium reaches HW equilibrium in a single generation of random mating
- ❖ differences between sexes in allele frequencies at autosomal loci are eliminated in a single generation of random mating

Statistical Tests

- ❖ Chi-squared test
 - ❖ estimate allele frequencies from phenotypes or directly from genotypes
 - ❖ use to calculate “expected” genotype frequencies
 - ❖ compare with observed genotype frequencies

$$\chi^2 = \sum_{i=1}^k \frac{(E_i - O_i)^2}{E_i} = \sum_{i=1}^k \frac{(\text{observed}_i - \text{expected}_i)^2}{\text{expected}_i}$$

- ❖ What does a “statistically significant” result tell us?

<http://faculty.vassar.edu/lowry/csfit.html>

Dominance

- ❖ with dominance and allele frequencies estimated from **phenotypes**, there are no degrees of freedom left for a statistical test
- ❖ $f_a = \hat{q} = \sqrt{R}$, where R is the frequency of the recessive phenotype

Other Statistical Tests

- ❖ Chi-squared test not valid for small sample sizes, corrections available but may be too conservative

- ❖ Alternatives:

- ❖ Exact test

$$\Pr\{n_{12} \mid n_1, n_2\} = \frac{n! / (n_{11}! n_{12}! n_{22}!)}{(2n)! / (n_1! n_2!)} 2^{n_{12}}$$

- ❖ Permutation test

Multiple tests

- ❖ Testing multiple loci for HWE increases the likelihood of false positives

- ❖ Remedies:

- ❖ Bonferroni correction: $p < 0.05/m$, where m is the number of independent tests

- ❖ much too conservative!

- ❖ Control of “false discovery rate”

- ❖ e.g., Benjamini & Hochberg (1995, J. Roy. Stat. Soc.)

- ❖ order p -values from smallest to largest, then find the largest i for which:

$$P_i \leq \frac{i}{m} q^*$$

Additional points...

- ❖ rarity of homozygotes for rare allele
 - ❖ most copies of rare alleles (e.g., recessive deleterious alleles) are in heterozygotes
- ❖ X-linked genes
 - ❖ male genotypes = allele frequencies
 - ❖ in XY systems
 - ❖ X-linked diseases
- ❖ multiple alleles
 - ❖ frequency of A_iA_i homozygotes = f_i^2
 - ❖ frequency of A_iA_j heterozygotes = $2f_i f_j$

Deviations from HWE

- ❖ Difference in allele frequencies between males and females
 - ❖ takes 2 generations to achieve HWE
- ❖ Assortative Mating or Dis-assortative Mating
- ❖ Inbreeding
 - ❖ similar effect to assortative mating
 - ❖ but, affects all loci across the genome
- ❖ Population Structure
 - ❖ different allele frequencies in different populations
- ❖ Selection
 - ❖ requires "strong" selection plus large sample size to detect the effect

The Inbreeding Coefficient

- ❖ Measures the reduction in heterozygosity in a population due to inbreeding

$$F = \frac{(2f_A f_a - f_{Aa})}{2f_A f_a}$$

- ❖ Rearranged, you can predict the frequency of heterozygotes given allele frequencies and F

$$F_{Aa} = 2f_A f_a (1 - F)$$

Multi-locus Genotypes

- ❖ The “joint probability” of genotypes at two or more independent loci is the product of their individual probabilities
- ❖ The basis for most forensic genetics
- ❖ E.g., ...