2. Hardy–Weinberg Equilibrium (Ch. 2)

Assumptions: Infinite size, random mating, no selection/mutation/migration. Two alleles A, a with $p, q = 1 - p$

Genotype Frequency: $P(AA) = p^2, \quad P(Aa) = 2pq, \quad P(aa) = q^2$

Allele Frequency: $p = P_{AA} + \frac{1}{2}P_{Aa}, \quad q = 1 - p$

Rare recessive disease: If prevalence $x = P(aa)$, then $q = \sqrt{x}$. Carrier freq $\approx 2q$ when $q \ll 1$.

Multiple alleles (K): $P(A_iA_i) = p_i^2, \quad P(A_iA_j) = 2p_ip_j \ (i \neq j)$

Goodness-of-fit test: $X^2 = \sum \frac{(O_i - E_i)^2}{E_i}, \quad df = (\#\text{genotypes} - 1) - (\#\text{alleles} - 1)$

Why HWE fails: Wahlund effect, inbreeding/assortative mating, selection, mutation/migration.

3. Linkage Disequilibrium (Ch. 3)

Recombination fraction: $\theta = g_{Ab} + g_{aB}$, with $0 \leq \theta \leq 0.5$. If $\theta < 0.5 \rightarrow$ linked.

LD parameter: $D = g_{AB} - p_Ap_B = g_{AB}g_{ab} - g_{Ab}g_{aB}$

Gamete frequencies: $g_{AB} = p_Ap_B + D$, $g_{Ab} = p_Ap_b - D$, $g_{aB} = p_ap_B - D$, $g_{ab} = p_ap_b + D$

LD decay: $D_n = D_0(1 - \theta)^n$. Next gen: $g^*_{AB} = g_{AB} - \theta D$

Normalized LD (-1 < D' < 1): $D' = D/D_{max}$ where $D_{max} = \min(p_Ap_b, p_ap_B)$ if $D \geq 0$, $D_{max} = \min(p_Ap_B, p_ap_b)$ if $D \leq 0$

An alternative measure Correlation: $r^2 = D^2/(p_Ap_ap_Bp_b)$

Gamete freq estimation: $g_{AB} = \frac{2N_{AABB} + N_{AaBB} + N_{AABb}}{2(N_{total} - N_{AaBb})}$

Test $H_0 : D = 0$: $\chi^2 = \frac{ND^2}{p_Ap_ap_Bp_b} \sim \chi^2_1$

4. Quantitative Genetics (Ch. 4)

Trait model: $P = G + E$, with $G = A + D + I$ (additive, dominance, epistasis)

Single-locus scaling: $A_1A_1 \rightarrow +a$, $A_2A_2 \rightarrow -a$, $A_1A_2 \rightarrow d$. Population mean: $\mu = a(p - q) + 2pqd$

Average effect: $\alpha = a + d(q - p)$; $\alpha_1 = q\alpha$, $\alpha_2 = -p\alpha$

Breeding values: $A_{11} = 2q\alpha$, $A_{12} = (q - p)\alpha$, $A_{22} = -2p\alpha$

Dominance deviations: $D_{11} = -2q^2d$, $D_{12} = 2pqd$, $D_{22} = -2p^2d$. Note: $G = A + D$

Variance components: $V_P = V_G + V_E = V_A + V_D + V_E$

Additive variance: $V_A = 2pq\alpha^2 = 2pq[a + d(q - p)]^2$. **Dominance**: $V_D = (2pqd)^2$

Heritability: Broad-sense $H^2 = V_G/V_P$; Narrow-sense $h^2 = V_A/V_P$

Covariance of relatives: $\text{Cov}(G_X, G_Y) = \frac{r_1}{2}V_A + r_2V_D$ where r_1=coeff. of relationship, r_2=prob. IBD at both alleles

Relationship	Covariance	h^2 estimation
Parent-offspring	$\frac{1}{2}V_A$	$h^2 = 2b_{OP}$
Mid-parent	$\frac{1}{2}V_A$	$h^2 = b_{O\bar{P}}$
Full sibs	$\frac{1}{2}V_A + \frac{1}{4}V_D$	
Half sibs	$\frac{1}{4}V_A$	$h^2 = 4b$

Offspring prediction: Expected deviation $= h^2x$ (mid-parent) or $\frac{1}{2}h^2x$ (single parent)

5. Natural Selection (Ch. 5)

Fitness: Relative fitness $w_{ij} = W_{ij}/W_{std}$. Mean fitness: $\bar{w} = P_{AA}w_{AA} + P_{Aa}w_{Aa} + P_{aa}w_{aa}$

Freq. after selection: $p^* = \frac{P_{AA}w_{AA} + \frac{1}{2}P_{Aa}w_{Aa}}{\bar{w}}$

Change in freq (HWE): $\Delta p = \frac{p}{\bar{w}}(w_A - \bar{w}) = \frac{pq}{\bar{w}}(w_A - w_a)$

Avg. allele fitness: $w_A = \frac{2w_{AA}P_{AA} + w_{Aa}P_{Aa}}{2P_{AA} + P_{Aa}} = pw_{AA} + qw_{Aa}$ (HWE);

Average excess: $a_A = w_A - \bar{w}$. Evolution $\propto$ avg excess, not genotype fitness

Selection coefficients: $w_{AA} = 1$, $w_{Aa} = 1 - hs$, $w_{aa} = 1 - s$. $h = 0$: recessive; $h = 1$: dominant; $h = 0.5$: additive

Equilibrium (set $w_{Aa} = 1$, $w_{AA} = 1 - s$, $w_{aa} = 1 - t$): $\boxed{\tilde{p} = \frac{t}{s + t}}$, $\tilde{q} = \frac{s}{s+t}$

Overdominance ($w_{Aa} > w_{AA}, w_{aa}$): Stable equilibrium $\rightarrow$ balanced polymorphism

Underdominance ($w_{Aa} < w_{AA}, w_{aa}$): Unstable equilibrium

Directional selection: One allele fixed $\rightarrow$ transient polymorphism

Fisher's FTNS: $\frac{d\bar{m}}{dt} = V_A$ (rate of $\uparrow$ mean fitness = additive genetic variance)

Wright's equation: $\Delta p = \frac{p(1-p)}{2\bar{w}}\frac{d\bar{w}}{dp}$

6. Mutation and Genetic Variation (Ch. 6)

Recurrent Mutation: $\Delta p = vq - up = -p(u + v) + v \Rightarrow p_e = \frac{v}{u+v}, \quad q_e = \frac{u}{u+v}$

Infinite-Alleles Model: Each mutation creates unique allele.

Infinite-Sites: At most 2 alleles/site (SNPs).

Mutation Rates: Per locus $\sim 10^{-5} - 10^{-6}$/gen; microsatellites $\sim 10^{-3}$; per nucleotide $\sim 10^{-8} - 10^{-9}$/gen.

Polymorphism: $P = \frac{\#\text{ polymorphic loci}}{\text{total loci}}$

Heterozygosity: $H_o =$ observed; $H_E = 1 - \sum p_i^2$; unbiased: $\hat{H}_E = \frac{2n}{2n-1}[1 - \sum \hat{p}_i^2]$

Nucleotide Polymorphism: $S_n^* = S_n/B$ ($S_n =$ segregating sites, $B =$ total nucleotides)

Watterson's Estimator: $\hat{\theta}_W = \frac{S_n}{\sum_{i=1}^{n-1}\frac{1}{i}}$ where $\theta = 4N_e\mu$

Nucleotide Diversity: $\hat{\theta}_\pi = \frac{\sum_{i<j} T_{ij}}{\binom{n}{2}}$. At equilibrium: $\theta_W = \theta_\pi$

7. Inbreeding (Ch. 7)

IBS: Two alleles functionally equivalent. **IBD**: Derived from same ancestor. IBD $\Rightarrow$ IBS.

Pedigree inbreeding coef. F_X: Prob. homozygous due to IBD.

Co-ancestry Φ_{XY}: Prob. random alleles from X,Y are IBD.

$\Phi_{XX} = \frac{1}{2}(1+F_X)$, $\Phi_{XY} = F_{X\times Y}$ (offspring's F)

Path method: $F = \sum_i \left(\frac{1}{2}\right)^{n_i}(1+F_{a_i})$

where n_i = # in path, F_{a_i} = ancestor's F. Note: $0 \le F \le 1$

Inbreeding Depression: $\ln S = -A - BF$. Humans: $5 < 2B < 8$ lethal equivalents.

System-of-mating inbreeding coef. f:

$f = corr(X_A, Y_A) = \frac{P_{AA}-p^2}{pq} = 1 - \frac{H_0}{H_{exp}} = 1 - \frac{P_{Aa}}{2pq}$. Note: $-1 \le f \le 1$

Genotypes with f: $P_{AA} = p^2 + pqf$, $P_{Aa} = 2pq(1-f)$, $P_{aa} = q^2 + pqf$

Recessive disease ratio: $\frac{q^2+pqf}{q^2} = 1 + \frac{pf}{q} \approx 1 + (\frac{1}{q} - 1)f$

8. Genetic Drift and Coalescence (Ch. 8)

Wright-Fisher: Diploid, constant N, random mating, no mutation/selection.

$X_{t+1}|X_t \sim \text{Bin}(2N, X_t/2N)$ $\text{var}(p_{t+1}|p_t) = \frac{p_t(1-p_t)}{2N}$

Fixation: Pr(fixed) $= p_0$; Pr(lost) $= 1-p_0$, not related to population size

Variance over time: $\text{var}(p_t) = p_0(1-p_0)[1-(1-\frac{1}{2N})^t]$

IBD increase: $\bar{F}_t = \frac{1}{2N} + (1-\frac{1}{2N})\bar{F}_{t-1}$

Heterozygosity decay: $\bar{H}_t = \boxed{(1-\frac{1}{2N})^t \bar{H}_0} \approx \bar{H}_0 e^{-t/(2N)}$

Founder effect: Pr(allele lost) $= (1-p)^{2N_f}$; $\text{var}(p_f) = \frac{p_a(1-p_a)}{2N_f}$

New mutation loss: Family size $\sim$ Poisson(2), Pr(loss in 1 gen) $= e^{-1} \approx 0.368$

Poisson Distribution: $X \sim \text{Pois}(\lambda)$: $\Pr(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$, $E[X] = \lambda$, $\text{Var}(X) = \lambda$

Effective size:

- Observed ave. prob. IBD: $N_{ef} = \frac{1}{2\{1-(1-\bar{F}^*)^{1/t}\}}$ where $\bar{F}^* = 1-(1-\frac{1}{2N_{ef}})^t$
- Unequal sex: $N_e = \frac{4N_m N_f}{N_m+N_f}$;
- Variable population: $N_e \approx t/\sum(1/N_i)$ (harmonic mean)
- Variable allel freq.: $\text{var}^* p_t = p_0(1-p_0)\left\{1-\left(1-\frac{1}{2N_{ev,t}}\right)^t\right\}$

Humans: $N \approx 6.7\times 10^9$, $N_e \approx 10^4$

Drift-Mutation Balance: $\boxed{\bar{F}_{eq} = \frac{1}{1+\theta}}$ where $\theta = 4N_e\mu$

Coalescent: $P_n = \frac{n(n-1)}{4N_e}$ (prob. coalescence 1-gen back);

Coalescence time distribution : $\Pr\{T_n = t\} = (1-P_n)^{t-1}P_n$ (Geometric distribution) $E[T_j] = \frac{4N_e}{j(j-1)}$, $E[T_{MRCA}] = 4N_e(1-\frac{1}{n})$; for $n=2$: $E[T_{MRCA}] = 2N_e$, $E[S_n] = \theta\sum_{i=1}^{n-1}\frac{1}{i}$

Geometric Distribution: $X \sim \text{Geom}(p)$: $\Pr(X=k) = (1-p)^{k-1}p$, $E[X] = 1/p$, $\text{Var}(X) = (1-p)/p^2$

9. Gene Flow and Population Subdivision (Ch. 9)

Gene flow model: $p_1^* = (1-m)p_1 + mp_2$; $\Delta p_1 = -m(p_1-p_2)$; $d_t = (1-2m)^t d_0 \to 0$

IBD with gene flow: $\bar{F}_t = \left\{\frac{1}{2N_{ef}} + \left(1-\frac{1}{2N_{ef}}\right)\bar{F}_{t-1}\right\}(1-m)^2$

Flow-Drift Balance: $\boxed{F_{ST} \approx \frac{1}{1+4N_e m}}$. $N_e m > 1 \Rightarrow F_{ST} < 0.20$

Wahlund Effect: Lumping subpops $\Rightarrow$ heterozygote deficiency

$\bar{P}_{11} = \bar{p}^2 + \sigma_p^2$, $\bar{P}_{12} = 2\bar{p}\bar{q} - 2\sigma_p^2$, $\bar{P}_{22} = \bar{q}^2 + \sigma_p^2$, $\sigma_p^2 = var(p)$

Heterozygosity levels: H_I = observed het. (individuals); H_S = expected het. within subpops; $H_T = 2\bar{p}\bar{q}$ = expected het. in total pop.

F-statistics: $f_{IS} = \frac{H_S-H_I}{H_S}$; $f_{ST} = \frac{H_T-H_S}{H_T} = \frac{\sigma_p^2}{\bar{p}\bar{q}}$; $f_{IT} = \frac{H_T-H_I}{H_T}$

f_{IS} = within-subpop inbreeding; $f_{IT} = f_{ST} + (1-f_{ST})f_{IS}$

$(1-f_{IT}) = (1-f_{IS})(1-f_{ST})$

Genotypes: $\bar{P}_{11} = \bar{p}^2 + \bar{p}\bar{q}f_{IT}$, $\bar{P}_{12} = 2\bar{p}\bar{q}(1-f_{IT})$

Wright's f_{ST} **guidelines**: 0–0.05 little; 0.05–0.15 moderate; 0.15–0.25 great; \$>\$0.25 very great. Human ≈ 0.156

10. Selection Interactions (Ch. 10)

General: $\Delta p = \frac{p}{\bar{w}}\alpha_A + \Delta p_{other}$; Equilibrium: $\tilde{p} = -\bar{w}\frac{\Delta p_{other}}{\alpha_A}$

Selection-Mutation (fitness: $w_{AA}=1$, $w_{Aa}=1-hs$, $w_{aa}=1-s$):

Recessive ($h=0$): $\boxed{\tilde{q} = \sqrt{u/s}}$. Partial dominance ($h>0$): $\boxed{\tilde{q} = u/(hs)}$

Selection-Mutation-Inbreeding ($h=0$): $\boxed{\tilde{q} = u/(fs)}$

Example: $u = 10^{-6}$, $s=1$: Random ($f=0$): $\tilde{q} = 10^{-3}$; With $f=0.01$: $\tilde{q} = 10^{-4}$

Quick Reference Table

Topic	Formula
HWE genotypes	$p^2, 2pq, q^2$
LD decay	$D_n = D_0(1-\theta)^n$
LD test	$\chi^2 = ND^2/(p_A p_a p_B p_b)$
Additive variance	$V_A = 2pq\alpha^2$
Heritability	$h^2 = V_A/V_P = 2b_{OP}$
Mutation equilibrium	$q_e = u/(u+v)$
Expected het	$H_E = 1 - \sum p_i^2$
Watterson's θ	$\hat{\theta}_W = S_n/\sum(1/i)$
Pedigree F	$F = \sum(1/2)^{n_i}(1+F_{a_i})$
Inbreeding genotypes	$P_{AA} = p^2 + pqf$
Drift variance	$\text{var}(p\|p_t) = p_t(1-p_t)/(2N)$
Drift-mutation	$\bar{F} = 1/(1+\theta)$, $\theta = 4N_e\mu$
Coalescence time	$E[T_j] = 4N_e/[j(j-1)]$
Effective size (sex)	$N_e = 4N_m N_f/(N_m+N_f)$
Variance eff. size	$N_{ev} \approx t/\sum(1/N_i)$
Gene flow-drift	$F_{ST} = 1/(1+4N_e m)$
F-statistics	$(1-f_{IT}) = (1-f_{IS})(1-f_{ST})$
Selection-mutation	$\tilde{q} = \sqrt{u/s}$ (recessive)
Sel-mut-inbreeding	$\tilde{q} = u/(fs)$