Serik Sagitov, Chalmers Tekniska Högskola, January 17, 2006

Population genetics course

Three major areas of genetics

Classical genetics

Mendel's principles; chromosomal mapping

Molecular genetics

DNA structure; transcription and translation

Evolutionary genetics

population genetics: gene frequencies

quantitative genetics: heritability of traits

phylogenetics: gene trees and species trees

Genetic terminology

DNA = deoxyribonucleic acid, two strands form a double-helix

four letters = nucleotides A, C, G, T

A binds to T and G binds to C

purines A,G and pyrimidines T,C

Human nuclear genome 3 000 000 000 base pairs

mitochondrial genome 16 000 base pairs

RNA = ribonucleic acid

one strand looped, letters A, C, G, U

Proteins

twenty letters = twenty amino acids

Protein synthesis, transcription and translation:

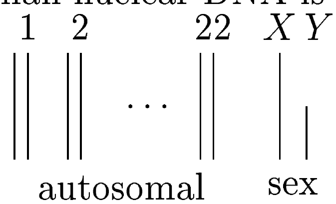
gene (a piece of DNA) $\to$ RNA $\to$ protein

Genetic code is degenerate, Table 1.1, p. 7

one codon (3 nucleotides) $\to$ one amino acid

61 codons $\to$ 20 amino acids, 3 codons $\to$ stop, $4^3 = 64$

Human nuclear DNA is packed in 23 pairs of chromosomes



XX = female
XY = male

Chromosome assortment

mother $(M_1^1 F_1^1 | M_2^1 F_2^1 | \ldots | M_{22}^1 F_{22}^1 | M_X^1 F_X^1)$
father $(M_1^2 F_1^2 | M_2^2 F_2^2 | \ldots | M_{22}^2 F_{22}^2 | M_X^2 Y)$
after meiosis and recombination
gametes $(M_1 | M_2 | \ldots | M_{22} | M_X)$ and $(F_1 | F_2 | \ldots | F_{22} | F_X)$
after mating
daughter $(M_1 F_1 | M_2 F_2 | \ldots | M_{22} F_{22} | M_X F_X)$

Alleles: different variants of a gene
gene A with alleles (A, a), gene B with alleles (B, b)

One locus genotypes
homozygous AA, aa; heterozygous Aa

Two loci genotypes
$\frac{AB}{AB}, \frac{AB}{Ab}, \frac{AB}{aB}, \frac{AB}{ab}, \frac{Ab}{Ab}, \frac{Ab}{aB}, \frac{Ab}{ab}, \frac{aB}{aB}, \frac{aB}{ab}, \frac{ab}{ab}$

Phenotype = an observable trait of an organism
codominant alleles: AA, Aa, aa look different

Dominant allele A, recessive a
if AA and Aa look similar, while aa look different

Course content

1. HWE and inbreeding coefficient
2. Mutation, migration, and selection
3. Random genetic drift
4. Molecular population genetics
5. Quantitative genetics

1. HWE and inbreeding coefficient

1.1 Genetic variation

Two measures of genetic variation

Polymorphism = proportion of polymorphic genes with most common allele frequency $p \leq 0.95$

Heterozygosity = proportion of heterozygous genes in an average individual

Ex 1: numerical example

Next table gives an example of a sample of four individuals with Pm = 0.3, and $\bar{H} = 0.1$

Assignment

1) explain the meaning of the ratio $\frac{\bar{H}}{\text{Pm}}$
2) using the same format suggest two other samples with Pm = 0.1, $\bar{H} = 0.1$ and Pm = 1.0, $\bar{H} = 0$

Genes	1*	2	3*	4	5	6	7	8	9	10*	$\bar{H}_{\text{ind}}$
Ind. 1	+	+	+	+	+	+	+	+	+	+	
	−	+	+	+	+	+	+	+	+	+	0.1
Ind. 2	+	+	−	+	+	+	+	+	+	+	
	+	+	+	+	+	+	+	+	+	+	0.1
Ind. 3	−	+	+	+	+	+	+	+	+	+	
	+	+	+	+	+	+	+	+	+	−	0.2
Ind. 4	+	+	−	+	+	+	+	+	+	+	
	+	+	−	+	+	+	+	+	+	+	0
$\hat{H}$	0.5	0	0.25	0	0	0	0	0	0	0.25	$\bar{H} = 0.1$

Ex 2: allozyme polymorphisms

Fig 2.9, p. 55: 14 to 71 genes (mostly ≈ 20) in 243 species

overall $\bar{x} \pm s$: Pm $= 0.26 \pm 0.15$, $\bar{H} = 0.07 \pm 0.05$

Drosophila species - most polymorphic group

mammals - least variable

cheetah almost monomorphic

$$\bar{x} := \frac{x_1 + \ldots + x_n}{n},\ s^2 := \frac{1}{n-1} \Sigma_{i=1}^{n} (x_i - \bar{x})^2$$

Ex 3: nuclear DNA polymorphisms

Alcohol dehydrogenase (*Adh*) in *D.melanogaster*

Fig 2.10, p. 58: 93 out of 113 alleles

Only two 2 allozymes due to a single

nonsynonimous mutation at amino acid number 193

slow allozyme *Adh-S*: AAG = Lysine,

fast allozyme *Adh-F*: ACG = Threonine

fast allele is more active and expressed

Ex 4: mtDNA polymorphisms

Fig 5.13, p. 188: 23 types of mtDNA

western-eastern subdivision of pocket gophers

Advantages with mtDNA analysis

higher mutation rate

maternal inheritance

no recombination

slow decomposition

1.2 Allele and genotype frequencies

one locus two allele model of a diploid population

Diploid population size N

genotype counts $N_{AA} + N_{Aa} + N_{aa} = N$

Haploid population size $2N$

allele counts $(2N_{AA} + N_{Aa}) + (2N_{aa} + N_{Aa}) = 2N$

Genotype frequencies

$D = \frac{N_{AA}}{N}$, $H = \frac{N_{Aa}}{N}$, $R = \frac{N_{aa}}{N}$

$D + H + R = 1$

Allele frequencies

$p = \frac{2N_{AA}+N_{Aa}}{2N} = D + \frac{H}{2}$, $q = \frac{2N_{aa}+N_{Aa}}{2N} = R + \frac{H}{2}$

$p + q = 1$

$$D = p^2 + pqF,\ R = q^2 + pqF,\ H = 2pq(1 - F)$$

$$\text{inbreeding coefficient } F = 1 - \frac{H}{2pq}$$

Sample frequencies

Sample counts in a random sample of n individuals

multinomial model: $(n_{AA}, n_{Aa}, n_{aa}) \in \mathrm{Mn}(n; D, H, R)$

Genotype frequencies and estimated standard errors

$\hat{D} = \frac{n_{AA}}{n}$, $\hat{H} = \frac{n_{Aa}}{n}$, $\hat{R} = \frac{n_{aa}}{n}$

$s_{\hat{D}} = \sqrt{\frac{\hat{D}(1-\hat{D})}{n-1}}$, $s_{\hat{H}} = \sqrt{\frac{\hat{H}(1-\hat{H})}{n-1}}$, $s_{\hat{R}} = \sqrt{\frac{\hat{R}(1-\hat{R})}{n-1}}$

Allele frequencies

$\hat{p} = \frac{2n_{AA}+n_{Aa}}{2n}$, $\hat{q} = \frac{2n_{aa}+n_{Aa}}{2n}$

$\mathrm{Var}(\hat{p}) = \frac{pq}{2n}(1+F)$, $s_{\hat{p}} = s_{\hat{q}} = \sqrt{\frac{\hat{p}\hat{q}}{2n}(1+\hat{F})}$, $\hat{F} = 1 - \frac{\hat{H}}{2\hat{p}\hat{q}}$

Ex 5: CCR5 gene

Human chemokine receptor gene

two alleles: A = no deletion, $a = \Delta 32$ deletion

genotype aa is resistant to HIV-1

Paris sample: $n = 294$, electrophoresis results

Band A (long)	—	—	
Band a (short)	—		—
Sample counts	64	224	6
Genotype	Aa	AA	aa

$\hat{D} = \frac{224}{294} = 0.76$, $\hat{H} = 0.22$, $\hat{R} = 0.02$

$s_{\hat{D}} = 0.025$, $s_{\hat{H}} = 0.024$, $s_{\hat{R}} = 0.008$

$\hat{p} = 0.87$, $\hat{q} = 0.13$, $\hat{F} = 0.03$, $s_{\hat{p}} = s_{\hat{q}} = 0.014$

Basques sample: $n = 111$, $\hat{q} = 0.018$, $s_{\hat{q}} = 0.009$

population founded 18000 years ago by a few imm.

Ex 6: RFLP

Restriction fragment length polymophisms
  due to restriction enzymes Fig 2.5, p. 49
Restriction enzyme EcoRI: restriction site GAATTC
  reveals an SNP like GAATTC → GATTTC
  since EcoRI can not cleave DNA
Southern blot procedure: Fig 2.7, p. 51
  allele a x—x——p—x
  allele A x————p—x
  x = restriction sites
  p = restriction site covered by radioactive DNA probe

Long fragment	—	—	
Intermediate	—		—
Short fragment	—	—	—
Sample counts	130	32	88

Southern blot results with $n = 250$
  $\hat{H} = 0.52$, $\hat{p} = 0.388$, $\hat{q} = 0.612$, $\hat{F} = -0.095$

1.3 Random mating and HWE

Dynamics of population frequencies over generations:
  $(D_0, H_0, R_0) \to (D_1, H_1, R_1) \to (D_2, H_2, R_2) \to \dots$
Hardy-Weinberg principle
  for given p_0 whatever are (D_0, H_0, R_0) we get
  $D_1 = p_0^2$, $H_1 = 2p_0q_0$, $R_1 = q_0^2$, $p_1 = p_0$, $q_1 = q_0$
  offspring inherit genes, not genotypes

H-W Equilibrium: $D = p^2$, $H = 2pq$, $R = q^2$

Hardy-Weinberg assumptions

1. diploid organisms
2. non-overlapping generations
3. effectively infinite population size $N = \infty$
4. random mating = panmixia
5. equal allele frequencies in the sexes
6. no mutation, 7. no selection, 8. no migration

Chi-square test of HWE

Test H_0: HWE using statistic $X^2 = \Sigma_{\text{cells}} \frac{(\text{obs}-\text{exp})^2}{\text{exp}}$

Asymptotic null distribution $X^2 \in \chi^2_{df}$

df = number of phenotypes – number of alleles

when df =1 use normal distribution table

Ex 6: RFLP

Expected (under HWE) genotype frequences

$\hat{D}_0 = \hat{p}^2 = 0.375$, $\hat{H}_0 = 2\hat{p}\hat{q} = 0.475$, $\hat{R}_0 = \hat{q}^2 = 0.150$

Cells	AA	Aa	aa	Total
Observed counts	88	130	32	$n = 250$
Expected counts	93.6	118.7	37.6	$n = 250$
$(\text{obs} - \text{exp})^2/\text{exp}$	0.335	1.076	0.834	$X^2 = 2.25$

P-value of the test: since df $= 3 - 2 = 1$

$\text{P}(X^2 \geq 2.25) = \text{P}(|\sqrt{X^2}| \geq 1.5)$

$\approx 2(1 - \Phi(1.5)) = 0.134$, accept H_0

Chi-square test and inbreeding coefficient: $X^2 = n \cdot \hat{F}^2$

Ex 5: CCR5 gene

Paris sample

$X^2 = 294 \cdot (0.03)^2 = 0.26$, df = 1, accept HWE

Estimation under HWE

Single gene recessive disease:

two phenotypes and two alleles, df $= 2 - 2 = 0$

cannot test HWE from phenotypes

Assuming HWE use estimate $\hat{q} = \sqrt{\hat{R}}$ with $s_{\hat{q}} = \sqrt{\frac{1-\hat{R}}{4n}}$

Ex 7: cystic fibrosis

CFTR gene, two alleles: normal A, mutant a

aa causes a severe condition, Caucasian $R = \frac{1}{2500}$

Assuming HWE for Caucasians

$q = \sqrt{R} = 0.02$ and $H = 2 \cdot 0.02 \cdot 0.98 = \frac{1}{26}$

Carriers to affected ratio $\frac{H}{R} = \frac{2p}{q} \approx \frac{2}{q}$

Propagation of error method

$f(\hat{R}) \approx f(R) + f'(R)(\hat{R} - R) + \frac{1}{2} f''(R)(\hat{R} - R)^2$

$\mathrm{E}(f(\hat{R})) \approx f(R) + \frac{1}{2} f''(R)\ \mathrm{Var}(\hat{R})$

If $f(x) = \sqrt{x}$, then $\mathrm{E}(\hat{q}) = \mathrm{E}(\sqrt{\hat{R}}) \approx \sqrt{R} - \frac{1}{8} R^{-3/2} \frac{R(1-R)}{n}$

$\mathrm{Var}(\hat{q}) = \mathrm{E}(\hat{R}) - (\mathrm{E}(\hat{q}))^2 \approx \frac{1-R}{4n}$

1.4 Inbreeding coefficient as correlation

Genotype $A_1 A_2$ sampled at random

$\mathrm{P}(A_1 = A_2 = A) = D$, $\mathrm{P}(A_1 = A_2 = a) = R$

$\mathrm{P}(A_1 = A, A_2 = a) = \mathrm{P}(A_1 = a, A_2 = A) = H/2$

$$\boxed{F = \text{correlation coeff. between } 1_{\{A_1=A\}} \text{ and } 1_{\{A_2=A\}}}$$

$F = 0$: independent alleles

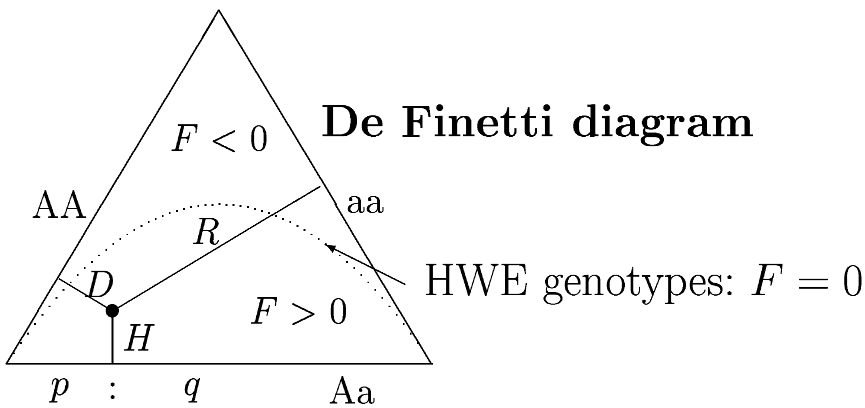
random genotype sampling = random allele sampling

$F > 0$: positive dependence

attraction of A to A and a to a, deficit of heterozygotes

$F < 0$: negative dependence

repulsion case, excess of heterozygotes



Ex 8: selfing

Mating genotypes: $AA \times AA$, $Aa \times Aa$, $aa \times aa$

$D_1 = D_0 + \frac{H_0}{4}$, $R_1 = R_0 + \frac{H_0}{4}$, $H_1 = \frac{H_0}{2}$

$D_t = p_0 - H_0 \cdot (0.5)^{t+1}$, $R_t = q_0 - H_0 \cdot (0.5)^{t+1}$

$H_t = H_0 \cdot (0.5)^t$, completely inbred line $F_t \to 1$

Ex 9: assortative mating

phenotype-based choice of mates: mating like-to-like for genes regulating the involved trait $F > 0$

Ex 10: disassortative mating

Mating to different phenotype: $(AA$ and $Aa) \times aa$

$D_1 = 0$, $R_1 = \frac{H_0}{2(D_0+H_0)}$, $H_1 = \frac{p_0}{D_0+H_0}$

$p_1 = \frac{H_1}{2}$, $F_1 = -\frac{H_1}{2-H_1}$

$D_2 = 0$, $R_2 = \frac{1}{2}$, $H_2 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $F_2 = -\frac{1}{3}$

which is the equilibrium distribution

Assortative mating effects certain genes
inbreeding effects the whole genome

1.5 HWE for multiple alleles

One locus with k alleles $A_1, A_2, A_3, \ldots, A_k$

genotype frequencies: $p_{11}, p_{12}, p_{13}, p_{23}, p_{33}, \ldots$

Number of possible genotypes

number of heterozygotes + number of homozygotes

$= \binom{k}{2} + k = \frac{k(k+1)}{2}$

Allele frequencies: $p_1, p_2, p_3, \ldots, p_k$

$p_i = p_i^2 + \frac{1}{2} \sum_{j \neq i} p_{ij}$

HWE genotype freqencies uniquely define p_i

A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3	$\ldots$
p_1^2	$2p_1p_2$	$2p_1p_3$	p_2^2	$2p_2p_3$	p_3^2	$\ldots$

$$\boxed{\text{HWE heterozygosity } H = 1 - p_1^2 - \ldots - p_k^2}$$

Ex 11: ABO blood groups

Three alleles and four phenotypes = blood groups

A = $\{AA, AO\}$, AB = $\{AB\}$

B = $\{BB, BO\}$, O = $\{OO\}$

Spanish Basques sample

Blood group	A	B	O	AB	Total
observed counts	724	110	763	20	n=1617
expected counts	710.7	94.8	776.12	35.4	n=1617

EM estimates of allele frequencies

$\hat{p}_A = 0.2661$, $\hat{p}_B = 0.0411$, $\hat{p}_O = 0.6928$

$X^2 = 9.58$, df $= 4 - 3 = 1$, $\sqrt{9.58} = 3.1$

reject HWE (possibly due to immigration)

Papago Indians, Arizona

Blood group	A	O	B	AB	Total
observed counts	37	563	0	0	n=600

Estimated allele frequencies under HWE

$\hat{p}_B = 0$, $\hat{p}_O = \sqrt{\frac{563}{600}} = 0.97$, $\hat{p}_A = 0.03$

different frequencies in two populations, why?

Ex 12: VNTR and DNA fingerprint

Variable number of tandem repeats, Fig 4.4, p. 130

minisatellite polymorphisms with 10-60 bp core repeat

Assuming 20 equally frequent alleles

$H = 1 - 20 \cdot (\frac{1}{20})^2 = 0.95$, Fig 4.5, p. 131

Evidence genotype (assumed to be heterozygous) against suspect genotype at n unlinked VNTR

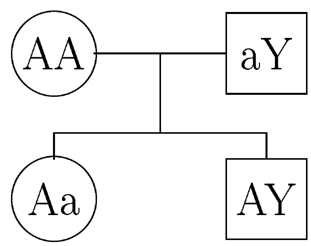
$P_n = \mathrm{P}(\text{perfect match})$, Fig 4.6, p. 133: $n = 9$

Several unlinked VNTR with 20 equally frequent alleles

$P_1 = 2 \cdot \frac{1}{20} \cdot \frac{1}{20} = \frac{1}{200}$, $P_n = P_1^n$

1.6 HWE for X-linked genes

One gene on the X chromosome, two alleles A and a



Allele A frequencies in males p_{m} and females p_{f}

dynamics of the frequencies: $p'_{\mathrm{m}} = p_{\mathrm{f}}$, $p'_f = \frac{p_{\mathrm{m}}+p_{\mathrm{f}}}{2}$

Equilibrium frequencies: $\hat{p}_{\mathrm{m}} = \hat{p}_{\mathrm{f}} = \frac{p_{\mathrm{m}}+2p_{\mathrm{f}}}{3}$

$$\boxed{\text{HWE: } D_{\mathrm{f}} = p^2,\ H_{\mathrm{f}} = 2pq,\ R_{\mathrm{f}} = q^2,\ p_{\mathrm{m}} = p_{\mathrm{f}} = p}$$

Recessive X-linked traits

affected males to females ratio $q_{\mathrm{m}}/R_{\mathrm{f}} = q/q^2 = 1/q$

Ex 13: color blindness

green blindness: $q = 0.05$, red blindness: $q = 0.01$

affected males to females ratios: 20 and 100

Ex 14: Xg blood group

X-linked gene with two alleles: $A = \text{Xg}^a$ and $a = \text{Xg}$

two blood types	Xg(a+)	Xg(a–)
female genotypes	Xg^a/Xg^a, Xg^a/Xg	Xg/Xg
male genotypes	Xg^a/Y	Xg/Y

British sample: female counts || male counts

	Xg(a+)	Xg(a–)	Total	Xg(a+)	Xg(a–)	Total
obs	967	102	1069	667	346	1013
exp	956.1	112.9	1069	683.8	329.2	1013

EM estimates: $\hat{p} = 0.675$, $\hat{q} = 0.325$

$X^2 = 2.45$, df $= 4 - 2 - 1 = 1$, $\sqrt{2.45} = 1.57$

not significant P-value $= 0.12$, do not reject HWE

1.7 Linkage disequilibrium (LD)

Two genes with two alleles each: A, a and B, b

actual gamete frequencies (left) and

linkage equilibrium frequencies (right)

	B	b	Tot
A	P_{11}	P_{12}	p_1
a	P_{21}	P_{22}	p_2
Tot	q_1	q_2	1

	B	b	Tot
A	p_1q_1	p_1q_2	p_1
a	p_2q_1	p_2q_2	p_2
Tot	q_1	q_2	1

Measures of LD

$P_{11} = p_1q_1 + D$, $P_{12} = p_1q_2 - D$

$P_{21} = p_2q_1 - D$, $P_{22} = p_2q_2 + D$

Basic LD measure $D = P_{11}P_{22} - P_{12}P_{21} = \text{Cov}(1_A, 1_B)$
depends on allele frequencies difficult to interpret

Correlation coefficient $r = \frac{D}{\sqrt{p_1 p_2 q_1 q_2}}$, $\hat{r}^2 = \frac{X^2}{n}$

Normalized D

$D' = \frac{D}{D_{\max}}$ if $D > 0$, where $D_{\max} = \min(p_1 q_2, p_2 q_1)$

$D' = \frac{D}{D_{\min}}$ if $D < 0$, where $D_{\min} = -\min(p_1 q_1, p_2 q_2)$

Ex 15: MN and Ss blood groups

Two genes in chromosome 4: alleles (M, N) and (S, s)
British sample, 1000 ind, $n = 2000$ chromosomes

Observed gamete counts and frequencies

	S	s	Total
M	474	611	1085
N	172	773	915
Tot	616	1384	2000

	S	s	Total
M	0.237	0.305	0.542
N	0.071	0.387	0.458
Tot	0.308	0.692	1

Linkage equilibrium frequencies and counts

	S	s	Total
M	0.167	0.375	0.542
N	0.141	0.317	0.458
Tot	0.308	0.692	1

	S	s	Total
M	334.2	750.8	1085
N	281.8	633.2	915
Tot	616	1384	2000

Chi-square test of independence: $X^2 = 184.9$, df $= 1$

$\sqrt{184.9} = 13.6$, reject H_0 : linkage equilibrium

$\hat{D} = 0.070$, $\hat{r} = 0.304$, $\hat{D}' = \frac{0.07}{0.141} = 0.5$

Attainment of linkage equilibrium

Changing D over generations under H-W assumptions

Fig 3.9, p. 100: $D_0 \to D_1 \to D_2 \to \ldots \to 0$

$$\boxed{D_t = D_0(1-\rho)^t, \text{ where } \rho = \text{recombination fraction}}$$

Causes of LD

1. small ρ, chromosome inversion
2. small t, recent mutation
3. epistatic selection favoring some genotypes
4. effectively small ρ, excess of homozygotes

Ex 16: LD in plants

Two unlinked esterase genes in Barley

gametes	B_1D_1	B_1D_2	B_2D_1	B_2D_2
observed counts	1501	754	720	74
LE expected counts	1642.6	613.7	577.1	215.6

$X^2 = 172.7$, df $= 1$, $D = -0.046$, $D' = 0.66$

significant LD due to 99% self-fertilization

Haldane's recombination model

Number of crossovers between two loci u Morgans apart

$X_u \in \text{Pois}(u)$ [definition of 1 Morgan: $\text{E}(X_1) = 1$]

$\rho = \text{P}(X_u \text{ is odd}) = \frac{1}{2}(1 - e^{-2u})$, $\rho \approx u$ for small u

$\rho \approx 0.5$ for large u, independent assortment

Ex 17: an assignment

Given the two loci genotype frequencies

	AB	Ab	aB	ab
AB	3/32	6/32	2/32	2/32
Ab	–	3/32	2/32	2/32
aB	–	–	3/32	6/32
ab	–	–	–	3/32

is the population in HWE? in LE?

Hint: first verify that

gamete and one locus genotype frequencies are

	B	b
A	0.25	0.25
a	0.25	0.25

	A	a
A	12/32	8/32
a	–	12/32

	B	b
B	12/32	8/32
b	–	12/32

1.8 Inbreeding coefficient as probability

Two alleles are IBD if they are derived

from a single allele in an ancestral HWE population

For an individual genotype any locus is

either autozygous: two IBD alleles, probability P(IBD)

or allozygous: non IBD alleles, probability $1 - \text{P(IBD)}$

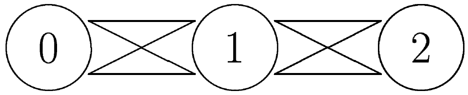
Pedigree formula of inbreeding coefficient
$F = \text{P(IBD)},\ F \geq 0$

$Fp = \text{P(autozygosity)} \times \text{P(ancestral allele is } A)$

$(1-F)p^2 = \text{P(allozygosity)} \times \text{P(ancestors are } A,\ A)$

$D = Fp + (1-F)p^2 = p^2 + pqF$

Ex 18: selfing



Forward time →

$1 - F_1 = \mathrm{P}(\overline{\mathrm{IBD}}) = \frac{1}{2}(1 - F_0)$, $1 - F_t = (\frac{1}{2})^t(1 - F_0)$

Complete inbreeding: $F_t \to 1$ as $t \to \infty$

One path with i ancestors $F_I = (\frac{1}{2})^i(1 + F_A)$

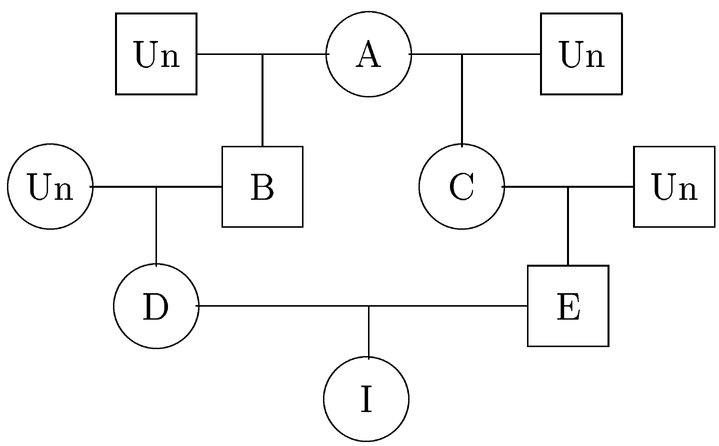
Ex 19: half-cousin mating

One path with five ancestors

$F_I = (\frac{1}{2})^5(1 + F_A)$

Half-cousin mating inbreeding coefficient

$F_I = 1/32$, if $F_A = 0$



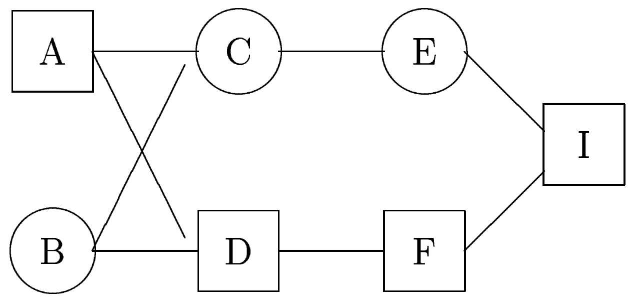
Ex 20: Speke's gazelle

St. Louis Zoo population founded with 1 male + 3 females

after 10 years: correlation $F = -0.333$

pedigree $F = 0.149$, close to half-sibs mating $F = 1/8$

Ex 21: first-cousin mating



Two mutually exclusive paths: FDACE and FDBCE

$F_I = (\frac{1}{2})^5(1 + F_A) + (\frac{1}{2})^5(1 + F_B)$

First-cousin mating inbreeding coefficient

$F_I = 1/16$, if $F_A = F_B = 0$

Ex 22: inbreeding depression

expression of hidden harmful recessives

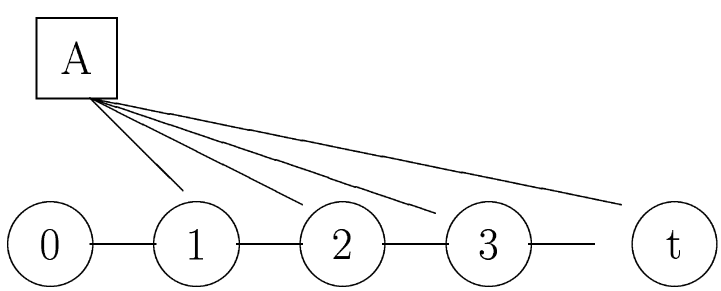
Rare recessive desease with $q = 0.01$:

random mating risk $q^2 = 0.0001$

first-cousin mating risk $R = q^2 + pqF = 0.0007$

Relative risk of a rare recessive desease $\frac{R}{q^2} \approx 1 + \frac{F}{q}$

Ex 23: repeated backcrossing



Autosomal gene: $t-1$ possible paths

$F_1 = 0$, $F_2 = \frac{1}{4}(1+F_A)$

$F_t = \frac{1}{4}(1+F_A) + \frac{1}{8}(1+F_A) + \ldots + (\frac{1}{2})^{t-1}(1+F_A)$

$F_t = (\frac{1}{2} - (\frac{1}{2})^t)(1+F_A) \to \frac{1+F_A}{2}$ as $t \to \infty$

Backcrossing to inbred strain: $F_A = 1$, $F_t \to 1$

backcrossing to random-bred strain: $F_A = 0$, $F_t \to \frac{1}{2}$

X-linked gene

$F_2 = \frac{1}{2}$, $F_3 = \frac{1}{2} + \frac{1}{4}$

$F_t = \frac{1}{2} + \frac{1}{4} + \ldots + (\frac{1}{2})^{t-1} = 1 - (\frac{1}{2})^t \to 1$

Fig 4.15, p. 154

pedigree F for different regular systems of mating

1.9 Inbreeding coefficient as fixation index

Metapopulation = K partially isolated HWE subpop-s

Diploid population sizes $N_i = w_i N$, $w_1 + \ldots + w_K = 1$

genotype frequencies $D_i = p_i^2$, $H_i = 2p_i q_i$, $R_i = q_i^2$

Metapopulation averages

$\bar{p} = \Sigma_{i=1}^K p_i w_i$

$D_S = \Sigma_{i=1}^K p_i^2 w_i = \overline{p^2}$, $H_S = 2\overline{pq}$, $R_S = \overline{q^2}$

Observed variance of allele freqs across subpopulations

$\sigma^2 = \overline{p^2} - (\bar{p})^2$

Complete allele fixation case: if $p_i = 0$ or 1, then

$\sigma^2 = \bar{p} - (\bar{p})^2 = \bar{p}\bar{q}$

Total population = hypothetical fused population with random mating

Expected genotype frequencies for the total population

$D_T = (\bar{p})^2$, $H_T = 2\bar{p}\bar{q}$, $R_T = (\bar{q})^2$

Wahlund's principle

isolation breaking increases genetic variation

$D_S - D_T = \sigma^2$, $R_S - R_T = \sigma^2$, $H_T - H_S = 2\sigma^2$

Isolation contributes to allele fixation

Fixation index $F_{ST} = 1 - \frac{H_S}{H_T} = \frac{\sigma^2}{\bar{p}\bar{q}}$

Inbreeding effect of population structure

$D_S = \bar{p}^2 + \bar{p}\bar{q}F_{ST}, R_S = \bar{q}^2 + \bar{p}\bar{q}F_{ST}$

$H_S = 2\bar{p}\bar{q}(1 - F_{ST})$

Ex 24: "desert snow" flowers

white flowers AA, Aa, blue flowers aa

Hierarchical structure: Fig 4.2, p. 114

metapopulation = three regions = 30 subpopulations

(West, Central, East) = (6, 20, 4) subpopulations

Table 4.1, p.115: average heterozygosities

observed $H_S = 0.1424$

expected assuming HWE regions $H_R = 0.1589$

expected under total HWE assumption $H_T = 0.2371$

$F_{SR} = 0.10$, $F_{RT} = 0.33$, $F_{ST} = 0.40$

Hierarchical formula $(1 - F_{ST}) = (1 - F_{SR})(1 - F_{RT})$
$F_{ST} \approx F_{SR} + F_{RT}$ for small F_{SR} and F_{RT}

Ex 25: codfish hemoglobin

Metapopulation sample

genotype	AA	Aa	aa	n
sample counts	130	763	1698	2591

Individual level average heterozygosity

$H_I = H = \frac{763}{2591} = 0.295$

Metapopulation level averages

$\bar{p} = 0.198$, $\bar{q} = 0.802$, $H_T = H_0 = 2\bar{p}\bar{q} = 0.317$

Overall inbreeding coefficient $F_{IT} = 1 - \frac{H_I}{H_T} = 1 - \frac{H}{H_0}$

$F_{IT} = 0.071$, $X^2 = 12.9$, df $= 1$, $\sqrt{12.9} = 3.6$

reject HWE hypothesis

Two races of cod recognized by anatomical differences

	AA	Aa	aa	n_i	p_i	H_i	F_i	$2p_iq_i$
Arctic	23	250	946	1219	0.1214	0.205	0.038	0.213
Coastal	107	513	752	1372	0.2649	0.374	0.041	0.390

Subpopulation level average heterozygosity

$H_S = 2\overline{pq} = 0.213 \cdot \frac{1219}{2591} + 0.390 \cdot \frac{1372}{2591} = 0.307$

Decomposition of the total inbreeding coefficient

fixation index $F_{ST} = 1 - \frac{H_S}{H_T} = 0.032$

inbreeding coefficient of mating $F_{IS} = 1 - \frac{H_I}{H_S} = 0.039$

Ex 26: three human subpopulations

Problem 4.4, p.126: compute pairwise fixation indices

gene	M	S	Fy^a	Jk^a	Js^a	β^s
blacks (Africa)	0.474	0.172	0	0.693	0.117	0.090
blacks (Georgia)	0.484	0.157	0.045	0.743	0.123	0.043
whites (Georgia)	0.507	0.279	0.422	0.536	0.002	0
F_{12}	10^{-4}	$4{\cdot}10^{-4}$	0.023	0.003	10^{-4}	0.009
F_{23}	0.001	0.016	0.268	0.026	0.059	0.047

MN blood groups data, 1 versus 2

$p_1 = 0.474$, $p_2 = 0.484$, $\bar{p}_{12} = 0.479$, $\bar{q}_{12} = 0.521$

$\sigma_{12}^2 = \frac{p_1^2+p_2^2}{2} - (\frac{p_1+p_2}{2})^2 = (\frac{p_1-p_2}{2})^2$

$F_{12} = \frac{(p_1-p_2)^2}{2\bar{p}_{12}\bar{q}_{12}} = 10^{-4}$

Duffy blood group

alleles Fy^a and Fy^b reveals very great differentiation between blacks and whites in Georgia

Fixation index scale

for the observed genetic differentiation

little differentiation:	$F_{ST} < 0.05$
moderate:	$0.05 \leq F_{ST} < 0.15$
great:	$0.15 \leq F_{ST} < 0.25$
very great:	$F_{ST} > 0.25$

Table 4.2, p. 121: fixation indices for various organisms