Population Genetics

Social Patterns and Evolutionary Forces in Human Populations

Population Genetics

- How do genes behave in populations
- What is a population?
  - A population is a subdivision of a species
  - A population is a community of individuals where mates are usually found
  - A population shares a common gene pool
  - A population has continuity through time

Allele frequency

- An allele frequency is the proportion of one allele relative to all alleles at the locus in the population
  - Example: In a population you draw blood samples and do the lab work to determine the hemoglobin type of the 50 individuals who comprise the population. You find:
    - 20 individuals with only type Hb^A
    - 20 individuals with both type Hb^A and type Hb^S
    - 10 individuals with only type Hb^S

Allele frequency, Example

- The frequency of the A allele is equal to the sum of all of the A alleles divided by the total number of hemoglobin alleles
  - = 40 A alleles (in 20 Hb^A/Hb^A individuals) + 20 A alleles (in 20 Hb^A/Hb^S individuals) divided by 100 total alleles
  - = 60/100
  - = **0.60 or 60%**
- S allele frequency
  - = 20 S alleles (in 20 Hb^S/Hb^S individuals) + 20 S alleles (in 20 Hb^A/Hb^S individuals) divided by 100 total alleles
  - = 40/100
  - = **0.40 or 40% <u>OR</u> 1 - freq(A)**

Allele frequency, Example cont.

- The frequencies of all alleles will always add up to 1 (or 100% of the alleles)
- The total number of alleles (not number of forms of the gene) for a given population at a given locus will be equal to two times the population size
  - Except for traits on the Y chromosome, where population size and allele count would be the same
- If there are only two alleles at the locus, there are only two allele frequencies and one of the frequencies will be equal to 1 - the frequency of the other allele

Genotype frequency

- A genotype frequency is the proportion of a population that has one genotype relative to all genotypes at a specific locus
  - In the previous example, we had 10 homozygous sicklers, genotype Hb^S / Hb^S, out of 50 individuals for a genotype frequency of 10/50 or one-fifth or 0.20 or 20%
  - A two allele locus will have three genotypes (except for Y-linked traits) and the frequencies of the three will add up to 1 or 100%

Hardy-Weinberg Equilibrium

- The Hardy-Weinberg model describes a mathematical relationship that allows the prediction of the frequency of offspring genotypes based on parental allele frequencies
- It also predicts that allele frequencies will not change from one generation to the next, i.e., it is an equilibrium or non-evolutionary model

Hardy-Weinberg Model

- In the parental generation of a population with a diallelic locus (alleles A and B), if the frequency of one allele (A) is p and the other allele (B) is q, i.e.,
  - frequency (A) = p,
  - frequency (B) = q,

  then the next generation will have:
  - frequency of the AA genotype = p^2
  - The frequency of the AB genotype = 2pq
  - The frequency of the BB genotype = q^2

Hardy-Weinberg Example

- At the MN blood group locus the frequency of the M allele equals 0.4 and the frequency of the N allele equals 0.6, the offspring in the next generation will have:
  - The frequency of the MM genotype = 0.16
  - The frequency of the MN genotype = 0.48
  - The frequency of the NN genotype = 0.36

Hardy-Weinberg Requirements

- **Random mating**
- No **mutation**
- Closed population, no **gene flow** (or migration of individuals) in or out
- Infinite size, no stochastic effects or **genetic drift**
- Equal fertility for all genotype groups--meaning no **selection** is occurring

Random Mating

- The H-W model requires that mating be random with regard to the locus being considered
  - The frequency of mating between males of one genotype and females of another should be equal to the product of the two genotype frequencies

Mating Types

Parents		Offspring		
Fathers	Mothers	AA	AB	BB
AA	AA	100%	--	--
AA	AB	50%	50%	--
AA	BB	--	100%	--
AB	AA	50%	50%	--
AB	AB	25%	50%	25%
AB	BB	--	50%	50%
BB	AA	--	100%	--
BB	AB	--	50%	50%
BB	BB	--	--	100%

Mating Types and Frequencies

Parents		Offspring		
Mating type	Freq	AA	AB	BB
AA x AA	.04	.04	.00	.00
AA x AB	.08	.04	.04	.00
AA x BB	.08	.00	.08	.00
AB x AA	.08	.04	.04	.00
AB x AB	.16	.04	.08	.04
AB x BB	.16	.00	.08	.08
BB x AA	.08	.00	.08	.00
BB x AB	.16	.00	.08	.08
BB x BB	.16	.00	.00	.16
Totals	**1.00**	**0.16**	**0.48**	**0.36**

Parental Genotype Frequencies:
Freq (AA) = 0.20, Freq (AB) = 0.40, Freq (BB) = 0.40

Shorthand H-W derivation

Females	Males	
	Freq(A) = p	Freq(B) = q
Freq(A) = p	Freq(AA) = p^2	Freq(AB) = pq
Freq(B) = q	Freq(AB) = pq	Freq(BB) = q^2

Random Mating

- If the frequency of the AA genotype in males is 0.2 and in females is also 0.2, then about 4% of all matings (0.2 x 0.2 = 0.04) should be between AA males and AA females
- If the frequency of mating is significantly different (test using χ^2) from the prediction, then there is significant deviation from random mating, and the H-W predictions for offspring genotype frequencies will be wrong

Assortative Mating

- If substantially more than the predicted frequency of matings are between males and females with the same genotypes, this would be an example of **positive assortative mating**
  - Positive assortative mating is the occurrence of mating between similar individuals at higher than random frequencies, resulting in more homozygotes than the H-W model predicts

Positive Assortative Mating

Parents		Offspring		
Mating type	Freq	AA	AB	BB
AA x AA	.20	.20	.00	.00
AA x AB	.00	.00	.00	.00
AA x BB	.00	.00	.00	.00
AB x AA	.00	.00	.00	.00
AB x AB	.40	.10	.20	.10
AB x BB	.00	.00	.00	.00
BB x AA	.00	.00	.00	.00
BB x AB	.00	.00	.00	.00
BB x BB	.40	.00	.00	.40
Totals	**1.00**	**0.30**	**0.20**	**0.50**

Parental Genotype Frequencies:
Freq (AA) = 0.20, Freq (AB) = 0.40, Freq (BB) = 0.40

Positive Assortative Mating

- As with most mammals, humans tend to mate with like individuals, particularly for visible or noticeable traits.

Trait	Spouse Correlation
I.Q.	**0.47**
Ear lobe length	**0.40**
Waist circumference	**0.38**
Height	**0.28**

Negative Assortative Mating

- If substantially fewer than the predicted frequency of matings are between males and females with the same genotypes, this would be an example of **negative assortative mating**
  - Negative assortative mating is the occurrence of mating between individuals with similar genotypes at lower than random frequencies, resulting in fewer homozygotes and more heterozygotes than the H-W model predicts

Negative Assortative Mating Example

- Negative assortative mating appears to be rare in mammals, but findings on rodents suggest that these mammals may have a preference for mates with dissimilar major-histocompatibility-complex (HLA) haplotypes
  - Ober and colleagues (1997) tested this mate preference in humans by surveying HLA haplotypes at five HLA loci among 411 Hutterite couples residing in 31 colonies in South Dakota

Negative Assortative Mating, 2

- Hutterites are a North American reproductive isolate originating in 1528 in the Tyrolean Alps
  - Approximately 400 members settled on three communal farms in South Dakota in the 1870s
  - There are now about 350 colonies and 35,000 individuals deriving from those original settlers
  - Marriage residence follows a patrilocal rule, while marriage may be either endogamous or exogamous with respect to the colony
- 41 couples matched for one haplotype, 2 matched for two haplotypes, and 1 man was homozygous for a haplotype matching one of his wife's haplotypes
  - 44/411 = 10.7% of couples matched for one or more haplotype

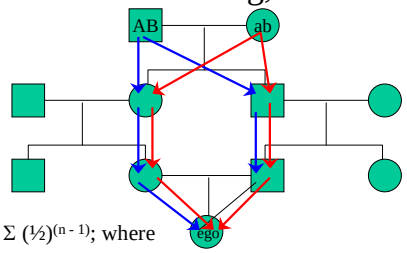
Negative Assortative Mating, 3

- The expected number of matches in 411 couples based on the assumption of random mating and the Hardy Weinberg model of the haplotype (allele) frequencies is 65/411 or 11/2%
  - The observed frequency is significantly lower expected (Chi-square test significant at p = 0.005)
- Variability in HLA haplotypes maximizes potential immune system response
- Mice detect HLA haplotype by smell of urine
- Humans may detect this through sweat odor
  - There is evidence that odor preferences may be HLA-linked in humans

Inbreeding

- Inbreeding, or mating between biologically related individuals at higher than random levels, increases homozygosity
  - Incest taboos prohibit mating between closely related individuals, making inbreeding less common than simple random mating would predict, and increasing heterozygosity
  - The inbreeding coefficient (F) is the probability of picking two alleles that are **identical by descent** (ibd) by a random draw in a population

Inbreeding, 2



$F = \Sigma\ (1/2)^{(n\ -\ 1)}$; where

n = number of links between an ancestor and ego, summed over all ancestor-ego loops

Above, n = 6, 2 loops, so $F = (1/2)^5 + (1/2)^5 = 1/16$, meaning there is a 1 in 16 chance of ego have alleles that are i.b.d.

Inbreeding, 3

- Small isolated populations end up with high levels of inbreeding, even when incest taboos are followed
  - Neel estimated that the average relationship (based on shared genes) between individuals in a Yanomamo village is nearly the same as between brothers and sisters
  - The result is increased homozygosity
  - Deleterious recessives show up more often

Deviations from Random Mating

- Assortative Mating and Inbreeding will both influence the relative frequencies of homozygotes and heterozygotes in the offspring generation
  - The frequency of alleles are NOT affected, unless some other forces are at work
  - The equilibrium prediction of unchanging allele frequencies are not affected by deviations from panmixia

Mutation

- Mutation is the alteration of the genetic material
  - Source of all **new** variability in the genome
  - Very small **quantitative** influence on allele and genotype frequencies
    - Mutation rate (μ) $\cong$ 0.00001 per generation per locus
      - Change allele frequencies by only about 1/100,000 per generation
  - Very significant **qualitative** impact on evolution through the genesis of unique new alleles, new forms of genes

Mutation, 2

- Since selection operates to optimize fitness, any give mutation is likely to reduce fitness
  - That is, any change to the coding sequence (exon portion, not introns) of a gene is likely to be detrimental and selected against

$$\mu = \frac{\text{Number of new mutations}}{\text{number of alleles in the population}}$$

- Example: Chondrodystrophic Dwarfism (Autosomal Dominant) D - dwarf; d - normal

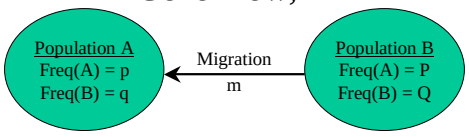
$$\mu = \frac{\text{\# Dwarfs born to normal parents}}{\text{2 times the number of individuals in the population}}$$

$= 79 \div 7{,}600{,}000 \cong$ **1/100,000**

Gene Flow, Migration

- Gene Flow is the intermarriage or genetic mixing between Mendelian populations
  - It has the effect of altering allele and genotype frequencies so that the two (or more) populations involved come to resemble each other in terms of genetic frequencies

Gene Flow, 2



After immigration, in population A:

sedente migrants

$q' = [(1 - m) \times q] + (m \times Q)$

$q' = q - mq + mQ$

Magnitude of change is determined by the allele frequency difference between the populations

$q' = q + [m \times \mathbf{(Q - q)}]$

Genetic Drift

- An infinite population size eliminates the chance or random influences on gene frequencies from one generation to the next which are especially significant in small populations
  - There are two primary manifestations of finite size and random fluctuations
    - Random Drift based on population size
    - Founder effect based on a random reduction in population size

Random Drift

- The genes of each generation are a random sample of preceding generations
  - The laws of probability apply to this sampling
    - Mean: The expected value of the allele frequency each generation is the same as the previous generation
      - $q_1 = q_0$
    - Standard deviation (σ) is a measure of dispersion about the mean, also an estimate of the probability of fluction from q_0
      - There is a 67% probability that q_1 will be within 1σ of q_0; a 95% probability within 2σ, and a 99% probability within 3σ

Random Drift, 2

- Example: Assuming a diallelic locus
  - freq(A) in generation 0 = p_0
  - freq(B) in generation 0 = q_0
  - Binomial distribution gives the following formula for the standard deviation of q_0:

$$\sigma_{q_0} = \sqrt{\frac{q_0 \times (1 - q_0)}{2 \times N_e}};$$

    - Where $p_0 = 1 - q_0$
    - and $2 \times N_e$ = 2 times the effective breeding population size (the number of alleles at the locus)

Founder Effect

- Founder Effect is the random fluctuation in allele frequencies caused by non-selection related reduction in population size followed by rapid population growth
  - The remaining population members become the random "founders" of the subsequent population
  - An example of the founder effect comes from the island of Tristan da Cunha, settled in 1816 by a group of 16 Scottish soldiers and their spouses

Tristan da Cunha

Year	Event	Size
1816	Settlement of island	16
1855	Dispute, causing 33 (of 103) to leave	70
1885	Population back up to 106, boat wreck kills 15 males, families begin to leave	106
1891	Population starts growing again	59
1961	Continued growth from 1891	270

Selection

- Selection causes changes in allele and genotype frequencies from one generation to the next due to differential net reproductive success of individuals with different genotypes
  - If individuals with genotype AA consistently have twice as many offspring as individuals with AB and BB genotypes, the frequency of the A allele will increase through time and eventually, everyone will have the AA genotype

Selection, 2

- There are two elements contributing to the differential reproductive success of individuals with differing genotypes
  - Viability or survival: individuals must survive to reproductive maturity in order to be able to reproduce
  - Fertility: individuals must produce offspring in order to pass on their genes on to the next generation of the population

Modeling Selection

- Selection operates by reducing the completed fertility of individuals with a certain phenotype, relative to other phenotypes within a population
  - Fitness ($1.0 \geq w \geq 0.0$) is the completed fertility for a given genotype, relative to the genotype with the highest completed fertility
  - The selection coefficient ($1.0 \geq s \geq 0.0$) measures the relative reduction in fertility for a genotype
    - $s = 1 - w$

Modes of Selection

Autosomal Inheritance	Genotypes		
Selection Against:	AA	AB	BB
Complete Dominance			
Dominant (Huntington's Disease)	s_{AA}	s_{AA}	
Recessive (PKU, Tay Sachs)			s_{BB}
Incomplete Dominance/Codominance			
Heterozygote		s_{AB}	
One homozygote, heterozygote	s_{AA}	s_{AB}	
Both homozygotes (sickle cell)	s_{AA}		s_{BB}

Modes of Selection, 2

Traits on X Chromosome	Females			Males	
Selection Against:	AA	AB	BB	A	B
Complete Dominance					
Dominant	s_{AA}	s_{AA}		s_{AA}	
Recessive (Hemophilia A)			s_{BB}		s_{BB}
Incomplete Dominance					
Heterozygote		s_{AB}			
1 homozygote, heterozygote	s_{AA}	s_{AB}		s_{AA}	
Both homozygotes (G6PD)	s_{AA}		s_{BB}	s_{AA}	s_{BB}
Traits on Y Chromosome					
Any allele (A)				s_{AA}	

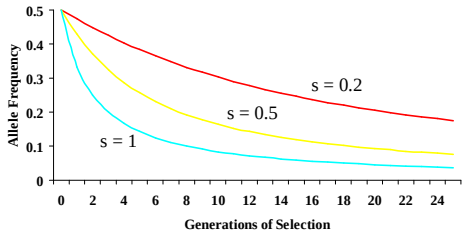
Selection against a recessive

	AA	AB	BB
Hardy-Weinberg Freq:	p^2	$2pq$	q^2
Fitness (w)	1.0	1.0	$1 - s_{(BB)}$
Frequency after selection	$p^2 \times (1)$	$2pq \times (1)$	$q^2 \times (1\text{-}s)$
New Total Frequency ($\bar{w}$) = $p^2 + 2pq + q^2 - q^2s = 1 - q^2s$			
Relative frequency after selection	$\frac{(p^2)}{(1 - q^2s)}$	$\frac{(2pq)}{(1 - q^2s)}$	$\frac{(q^2 - q^2s)}{(1 - q^2s)}$

Frequency of B in next generation = freq (BB) + ½ freq (AB)

$$q_1 = (q - q^2s) \div (1 - q^2s)$$

Selection against a recessive



Overdominance

	AA	AB	BB
Hardy-Weinberg Freq:	p^2	$2pq$	q^2
Fitness (w)	1.0	1.0	$1 - s_{(BB)}$
Frequency after selection	$p^2 \times (1-s_{AA})$	$2pq \times (1)$	$q^2 \times (1-s_{BB})$
New Total Frequency ($\bar{w}$) = $1 - p^2 s_{AA} - q^2 s_{BB}$			
Relative frequency after selection	$\frac{(p^2 - p^2 s_{AA})}{(\bar{w})}$	$\frac{(2pq)}{(\bar{w})}$	$\frac{(q^2 - q^2 s_{BB})}{(\bar{w})}$

Frequency of B in next generation = freq (BB) + ½ freq (AB)

$$q_1 = (q - q^2 s_{BB}) \div (1 - p^2 s_{AA} - q^2 s_{BB})$$

Equilibrium is attained if $\Delta q = 0$;

$\Delta q = q_1 - q_0 = [(q - q^2 s_{BB}) \div (1 - p^2 s_{AA} - q^2 s_{BB})] - q$

$= [pq \times (p s_{AA} - q s_{BB})] \div (1 - p^2 s_{AA} - q^2 s_{BB})$

$\Delta q = 0$ if $p s_{AA} - q s_{BB} = 0$; that is, if $q = s_{AA} \div (s_{AA} + s_{BB})$

Natural Selection

- Darwinian natural selection is a two-step process:
  - The production of new genetic variation through the process of mutation
  - The differential reproduction of favorable variants through the process of selection

Sources

- Ober, C.; Weitkamp, L.R.; Cox, N.; Dytch, H.; Kostyu, D.; Elias, S. 1997. HLA and mate choice in humans. American Journal of Human Genetics, 61:497-504.