Genotypes, Phenotypes and Hardy Weinberg Equilibrium

Biostatistics 666

Lecture II

Previously: Refresher on Genetics

- DNA sequence
  - Human Genome
- Inheritance of genetic information
- Sequence variation
  - VNTRs, microsatellites and SNPs

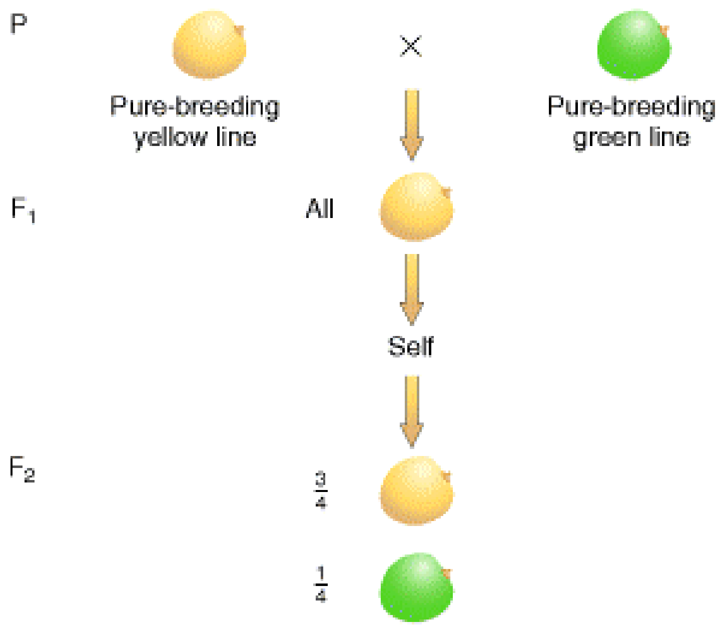
- Common Types of Genetic Study

Gregor Johann Mendel

- Discovered basic principles of genetics
  - “The Father of Genetics”
  - Monk, lived 1822 – 1884
- Crosses between strains of peas
  - Garden pea (*Pisum sativum*)
  - Each strain has particular characters
  - Height, flower color, seed shape …

Mendel's Experiment

- Crossed different true-breeding strains
  - Identical results for reciprocal crosses
- F_1 resembled one of the parental strains
- In F_2 generation, the other parental trait reappears …



Mendel's Numbers

- Seeds: Yellow vs. Green
  - F_1: All yellow
  - F_2: 6022 yellow, 2001 green
    - 75.1% yellow, 24.9% green
- Seeds: Smooth vs. Wrinkled
  - F_1: All smooth
  - F_2: 5474 smooth, 1850 wrinkled
    - 74.7% yellow, 25.3% wrinkled

Phenotype vs. Genotype

- **Genotype**
  - Underlying genetic constitution
- **Phenotype**
  - Observed manifestation of a phenotype

- The yellow peas in the parental and F_1 generations are not the same

Mendel’s Interpretation

- Each trait determined by “particulate factors” (**genes**)
  - E.g.: Seed colour
- Alternative forms for each factor (**alleles**)
  - E.g.: Yellow seeds or green seeds
- Each plant carried two alleles
  - Identical for true breeding parental strains
  - Different for F_1 generation

The Principle of Segregation

- **Mendel’s First Law**
- The two alleles of a gene pair segregate from each other in the formation of gametes

  (gametes are reproductive cells that fuse to form a new organism in sexual reproduction)

Genotypes

- Each individual carries two alleles
  - If there are **n** alternative alleles …
  - … there will be **n(n+1)/2** possible genotypes
- **Homozygotes**
  - The two alleles are the same
  - E.g.: Green/Green or Yellow/Yellow
- **Heterozygotes**
  - The two alleles are different
  - E.g.: Green/Yellow

Penetrance

- Describes the relationship between phenotypes and genotypes
  - **Complete Penetrance**
    - Each genotype corresponds to only one phenotype
  - **Incomplete Penetrance**
    - Link between phenotype and genotype is only probabilistic

The ABO blood group

- Important for blood transfusions
- Determined by alleles of the ABO gene
- 3 alternative alleles
  - A, B and O
- 6 possible genotypes, n (n + 1) / 2
  - A/A, A/B, A/O, B/B, B/O, O/O

ABO Blood Group II

Phenotype	Antigen		Antibody	
	A	B	A	B
A	+	-	-	+
B	-	+	+	-
O	-	-	+	+
AB	+	+	-	-

Relationships between alleles

- Relation between alleles
  - A and B are **dominant** over O
  - O is **recessive** in relation to A and B
  - A and B are **codominant**
- In this case all genotypes are fully **penetrant**

Genotype	Phenotype
A/A	A
A/B	AB
A/O	A
B/B	B
B/O	B
O/O	O

BRCA1 and Breast Cancer

- BRCA1 mutations predispose to breast cancer
  - About 0.1% of the population carries mutations in the BRCA1 gene
- Disease Risk

Age	40	60	80
Carriers	40%	70%	80%
Non-carriers	0.4%	3%	8%

Alleles, Genotypes and Phenotypes

- Classifying genotypes
  - Homozygous
  - Heterozygous
- Penetrance
- Relationships between alleles
  - Dominant, Recessive, Co-Dominant

Genes in Populations

- Genotype Frequencies
- Haplotype Frequencies
- Allele Frequencies
- Penetrance Function

- Derived measures of marker informativeness

Notation

- p_{ij}
  - frequency of genotype i/j in the population
  - $n(n+1)/2$ of these
- p_i
  - frequency of allele i in the gene pool
  - n of these
- Write allele frequencies as function of genotype frequencies

Hardy-Weinberg Equilibrium

- Random union of games
- Relationship discovered it in 1908
  - Hardy, British mathematician
  - Weinberg, German physician
- Shows **n** allele frequencies determine **n(n+1)/2** genotype frequencies
  - Large populations

Required Assumptions

- Diploid, sexual organism
  - Non-overlapping generations
- Autosome
- Large population
- Random mating
- Equal genotype frequencies among sexes
- Selection

Random Mating: Mating Type Frequencies

Mating	Frequency
$A_1A_1*A_1A_1$	p_{11}^2
$A_1A_1*A_1A_2$	$2p_{11}p_{12}$
$A_1A_1*A_2A_2$	$2p_{11}p_{22}$
$A_1A_2*A_1A_2$	p_{12}^2
$A_1A_2*A_2A_2$	$2p_{12}p_{22}$
$A_2A_2*A_2A_2$	p_{22}^2
Total	1.0

Mendelian Segregation: Offspring Genotype Frequencies

		Offspring		
Mating	**Frequency**	**A_1A_1**	**A_1A_2**	**A_2A_2**
$A_1A_1 * A_1A_1$	p_{11}^2	p_{11}^2		
$A_1A_1 * A_1A_2$	$2p_{11}p_{12}$	$p_{11}p_{12}$	$p_{11}p_{12}$	
$A_1A_1 * A_2A_2$	$2p_{11}p_{22}$		$2p_{11}p_{22}$	
$A_1A_2 * A_1A_2$	p_{12}^2	$\frac{1}{4}p_{12}^2$	$\frac{1}{2}p_{12}^2$	$\frac{1}{4}p_{12}^2$
$A_1A_2 * A_2A_2$	$2p_{12}p_{22}$		$p_{12}p_{22}$	$p_{12}p_{22}$
$A_2A_2 * A_2A_2$	p_{22}^2			p_{22}^2

And now...

$$
\begin{aligned}
p'_{11} &= p_{11}^2 + p_{11}p_{12} + \tfrac{1}{4}p_{12}^2 \\
&= (p_{11} + \tfrac{1}{2}p_{12})^2 \\
&= p_1^2 \\
p'_{22} &= p_{22}^2 + p_{22}p_{12} + \tfrac{1}{4}p_{12}^2 \\
&= (p_{22} + \tfrac{1}{2}p_{12})^2 \\
&= p_2^2 \\
p'_{12} &= 2p_{11}p_{22} + p_{11}p_{12} + p_{12}p_{22} + \tfrac{1}{2}p_{12}^2 \\
&= 2(p_{11} + \tfrac{1}{2}p_{12})(p_{22} + \tfrac{1}{2}p_{12}) \\
&= 2p_1p_2
\end{aligned}
$$

Conclusion

- Genotype frequencies are function of allele frequencies
  - Equilibrium reached in one generation
  - Independent of initial genotype frequencies
  - Random mating, etc. required
- Conform to binomial expansion
  - $(p_1 + p_2)^2 = p_1^2 + 2p_1p_2 + p_2^2$

A few more notes…

- Can be expanded to multiple alleles
  - Expand $(p_1 + p_2 + p_3 + \cdot \cdot \cdot \cdot + p_k)^2$
- Holds in almost all human populations
  - Little inbreeding (typical F = ~0.005)
- Deviations can suggest:
  - Problems with experimental assays
  - Non-independence of observations
  - Selection
  - Disease locus

Heterozigosity

- Probability that two alleles will differ

$$H = 1 - \sum p_i^2$$

- For *a* equally frequent alleles

$$H = 1 - \frac{1}{a} = \frac{a-1}{a}$$

- Sometimes called "gene diversity"

PIC

- Probability that alleles of parent can be distinguished in offspring
  - Botstein et al, 1980.
    - Markers that could track dominant alleles
- Probability that parent will heterozygous and informative in relation to spouse

PIC – Definition

- In general:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2(p_i p_j)^2$$

- For *a* equally frequent alleles

$$PIC = \frac{a-1}{a} - \frac{a-1}{a^3}$$

- PIC <= H

Exercise

- ABO locus allele frequencies
  - A – frequency 0.3
  - B – frequency 0.1
  - O – frequency 0.6
- Calculate genotype frequencies
- Calculate heterozygosity and PIC
- Calculate phenotype frequencies

NOD2 and Bowel Disease

- Leu1007fs
  - Frame shift mutation at position 1007
  - Frequency of about 5%
  - Disrupts gene
- Penetrance

Genotype	+/+	-/+	-/-
P(Crohn's|G)	0.1%	0.2%	3%

- Calculate frequency of -/- genotype in population and among patients…