POPULATION GENETICS - OVERVIEW

Prof. Arjumand S. Warsy
Department of Biochemistry, College of Science
King Saud University, Riyadh

Introduction

_ "Population genetics" is the study of the distribution of genes in populations and how these genes and genotype frequencies are maintained or changed.

_ Population genetics has much in common with "epidemiology" - the study of the interrelationships of the various genetic and environmental factors that determine the frequency and distribution of diseases in human communities.

_ "Genetic epidemiology" results from fusion of both 'population genetics' and 'epidemiology' and is concerned with diseases that have complex patterns of inheritance or are caused by a combination of heritable and environmental factors.

<u>The Human Population</u>

- Human population → 5 billion in the world

Human Population
↓
Sub-populations (races)
↓

Caucasians	Blacks	Asians
↓	↓ ↓	
Several genetically distinct sub-groups (ethnic groups)	Sub-groups (ethnic groups)	Sub-groups (ethnic groups)

- Races are major population groups whose gene pools differ from each other.
- Human chromosomes and loci that they contain are identical in all members, allele frequencies at many loci vary widely among population groups.
- Some variants are more frequent in one sub-group compared to others.
- Different alleles occur at different frequencies in different populations.
- Each ethnic group has its own characteristic sets of gene frequencies.
- Mutations are the basis of the genetic differences among races and among their sub-groups.

- How do these genetic differences become established in the different populations?
  - Selection of favourable mutations in response to environmental conditions
  - Chance survival of specific or even harmful mutations.
  - Reproductive isolation.

- Tables 1 & 2 present examples of disease alleles and normal polymorphic loci in some populations.

Maintenance of Gene Frequencies

- The relative frequencies of different alleles tend to be kept constant from one generation to the next.

- Consider one locus with two alleles A & α.
  Frequency of allele A = P
  Frequency of allele a = q
  The sum of these alleles must be 1 or 100% (P +q = 1).

- When the population is at genetic equilibrium, and when the genotype frequency of the offsprings is calculated, it is shown that the relative population of each genotype have remained constant (AA at P^2, Aa at 2pq and aa at q^2), i.e.

	A(P	a (q)	
A (p)	AA (P^2)	Aa (pq)	Frequency of each genotype as determined by the Punnetts
a (q)	Aa (pq)	aa (q^2)	Square.

$$P^2 + 2pq + q^2 = 1$$

- If there is random mating, each generation (1st, 2nd, 3rd and so on) will have the same genotype frequency. This principle is called the Hardy Weinburg.

Probability

Pobability is the ratio of the number of occurrence of a particular event to a total number of all possible events.
e.g. sex ratio in families:

- The probability that the first born child is a boy is 1/2. This probability 1/2 is repeated for every subsequent pregnancy and each child is an independent event. The probability that the first two children in a family will be boys is 1/2 x 1/2 = 1/4.
  When probabilities are applied to the distribution of boys and girls among all two child families, this follows the coincidence of a random event.

- Such probability of combination of two independent event, if extended is called "Binomial Distribution".

- If the probability of a boy is P(1/2) and the probability of a girl is q (1/2), then in a family (n = No. of children in the family) the frequencies of the different possible combination of P & q are presented by binomial distribution of (P + q). Thus for

two-child families

$(P + q)^2$ = $P^2 + 2Pq + q^2$
Families of 2 boys: $P^2 = 1/2 \times 1/2 = 1/4$
Families of 2 girls: $q^2 = 1/2 \times 1/2 = 1/4$
Families of 1 boy & 1 girl: $2Pq = 2 \times 1/2 \times 1/2 = 1/2$

The sex ratio is defined as the ratio of the number of male births to the number of female births.

Genes in populations

- The frequency of a hereditary disorder is independent of whether the disease is dominant or recessive.

- The Hardy-Weinberg Law put forward by a Cambridge mathematician Hardy, and a Stuttgart physician, Weinberg in 1908, is considered the corner stone of population genetics and it states that (neglecting mutations, selection, gene flow and genetic drifts) the gene frequency and genotype frequency remain constant from generation to generation.

$P^2 + 2Pq + 2q^2 = 1$
where P is the frequency of 'A' allele and q of 'a' allele.

Application of Hardy-Weinberg Law

Hardy-Weinberg Law is used for:

- Determining the gene frequencies when the genotype incidence is known. Gene frequencies:

  - help to compare the frequency of trait in different populations.
  - enable testing of the mode of inheritance of a given trait.
  - enable the calculation of the expected genotype frequencies.
  - enable the study of gene frequencies over time to see whether an equilibrium exist.

- Determining the carrier frequency for AR disorders from gene frequency.

- Determinig the proportion of heterozygotes when the frequency of the recessive phenotype is known.
  e.g. incidence of PKU (AR) is 1 in 10,000 live births

then $q^2 = 1/1000 = 0.0001$
$q = \sqrt{0.0001} = 0.01$
where q = gene frequency of PKU.
The frequency of heterozygotes is 2 Pq.
As P + q = 1.
P : 1 - 0.01 = 0.99 and

2 Pq: 2(0.99 x 0.01) = 0.0198 or 1/50.

1 person in 50 is a heterozygote or a carrier of PKU gene.

(Even though the persons affected by a rare recessive trait is only few, the frequency of the heterozygous carriers is relatively high).

Problem calculate the carrier frequency of the following AR disorders. The disease frequency is given

	Disease frequency
CF	1 in 2,500
Congenital deafness	1 in 5,000
Albinism	1 in 40,000
Alkaptonuria	1 in 100,000

- For X-linked recessive disorders

  The frequency is equal to the frequency of affected males, as males are homozygous:

  i.e. q = No. of affected males, and

  P = No. of unaffected males

  P + q = 1

  Females have 2 genotypes. q^2, P^2 and 2 Pq (like the autosomal inheritance). Males affected by XR disorder are more than females affected by this disorder.

For consanguineous mating

The risk of producing offspring homozygous for a certain recessive gene increases significantly in consanguineous mating. The closer the relationship of the parents the higher is the risk.

The number of genes common in two relatives is halved for each step further away in relationship.

For 1st cousins (3rd degree relatives) the coefficient of relationship (r) i.e. the chance that the two have a gene in common, or the proportion of all their genes which have been inherited from common ancestors is one-eight (1/8).

Genetic polymorphisms

- A genetic polymorphism is defined as "the occurrence together in a population of two or more genetically determined forms at such frequencies that the rarest of them could not be maintained by mutation alone".
- In general, a polymorphic loci is one in which there are at least two alleles, each with frequencies of greater than 1%.
- Alleles with frequencies less than 1% are said to be rare variants.
- Examples of polymophic loci:
  ABO blood group
  HLA
  Several plasma proteins (α1-AT, haptoglobin)
  Enzymes (G-6-PD)
  DNA variants (RFLPs); most genes)
- Polymorphism at DNA level have proved very valuable:

- in positional cloning studies leading to isolation of many disease genes
- in gene tracking, which have facilitated preclinical diagnosis, prenatal diagnosis, carrier detection.
- if the polymorphic markers has high "Polymorphic information content" (PIC) it is more likely that it will be of value in linkage analysis and gene tracking.

• Balanced and transient polymorphism
  - In 'balanced' polymorphism, two or more different forms are maintained by a balance between the selective advantage of the heterozygotes and the reduced fitness of the affected homozygotes e.g. Hb S in areas where malaria is endemic.
  - If the incidence of a particular polymorphic form decreases if the selective advantage is reduced, then such a situation is called "transcient" polymorphism.

<u>Genetic Linkage</u>

• According to Mendel's third Law: genes at different loci segregate independently. This is true for:
  - gene on different chromosomes
  - genes away from each other on the same chromosome

• But it is not true for loci positioned close to one another in the same chromosome. These are "syntenic' loci i.e. genes at these loci are inherited together and these loci are said to be 'linked'. These loci are so close that it is unlikely that they will be separated by a crossover or recombination during meiosis.

• Linked alleles on same chromosome are said to be in "coupling".

• Alleles on opposite chromosome are said to be in "repulsion".

• Recombinant fraction (θ) is a measure of the distance separating 2 loci.
If 2 loci are not linked $\theta = 0.5$ (i.e. the unlinked genes will segregate together during 50% of all meiosis).
$\theta = 0.05$ - syntenic allelles segregate together 19/20 times.

• Centimorgan (map unit)(cM): is a unit for measurement of genetic linkage.
If two loci are one CM apart, then a crossover occurs between them during on average 1/100 meiosis $\theta = 0.01$.
cM is not the same as the physical distance which is measure in bp e.g.

  - Human genome = 3000 cM in length.
    3×10^9 bp

  - Therefore IcM = 10^6 bp (1000 kb)

  - The relationship between linkage map units and physical length is not linear.

  - Some regions on chromosomes are more prone to recombinations, reason not known. These are called "hotspots".

  - Generally during meiosis:

1 large chromosome	=	on average has 3 crossover
1 medium "	=	on average has 2 crossover
1 small "	=	on average has 1 crossover

Linkage disequilibrium

- "The association of two linked alleles more frequently than would be expected by chance", is referred to as "linkage disequilibrium".

- e.g. if a disease allele is found to be in compling almost exclusively with one particular marker allele then they are said to be in linkage disequilibrium

- Linkage disequilibrium has been demonstrated in sickle cell disease, Huntington's disease, myotonic dystrophy, cystic fibrosis, phenylketonuria and fragile X syndrome.

Table 1: Frequency of disease alleles in different populations

Allele/Disease	Population differences
$ß^S$	↑ Africa, ↓ in Europe
$ß^C$	↑ in W. Africa, low elsewhere
CF	↑ in Europe & U.S., Caucasian ↓ in Asia and Africa and Finland
Tay-Sachs disease	↑ in Ashkenazi Jews, low in others
Myotonic dystrophy	~ in 10,000 - 20,000 in most populations ↑ in Quebec (<1 in 1000)

Table 2: Frequency of some polymorphic loci in different populations

Locus	Allelic variation
ABO Blood groups	Wide variations e.g. ↑ B- Asians ↓ in American
Other blood groups	Wide variations e.g. Rh allele R° found only in Africa
α1-AT	M1 0.51 - 0.98 M2 0.00 - 0.20
Alcohol dehydrogenase	ADH_2 - 90% in Japanese 15% in Europeans
Aldehyde dehydrogenase	$ALDH_1$ def. ~ 50% in Asians <50% in N. Americans