Genetics

Population Genetics

Prof. Dr. Akin Pala

http://akin.houseofpala.com

akinpala@gmail.com

Variation in Populations

Phenotypic Variation

- No two individuals are the same, even twins have some differences.
- However, one observation/individual does not inform us on variation.
  - Alien from space, green with antennas
- What is variation?
- Even a single trait has phenotypic variation
  - height, weight, survival, conception rate.

Phenotypic Variation

- Part of the phenotypic variation is determined by genes (genotype)
- These genes pass from generation to generation
- They form the basis for selection.

Quantitative-Qualitative

- Qualitative traits
  - Simple inheritance, Mendelian,
  - typically monogenetic, one or two genes
  - phenotype falls into different categories
  - Inherited diseases, blood type

Simply inherited traits

- Traits with simple inherit
several genes and their ge
categories.
- <u>Little</u> or no effect by envi
  - Coat color, eye color is not
  - Polled/horned condition
  - White man gains more pigm
does not become black
  - If a black person turns into a vampire, does he become white?

Quantitative-Qualitative

- Quantitative traits
  - Complex inheritance, polygenic
  - Continuous variation because the trait is the sum of several small effects caused by many genes
  - There are QTL (Quantitative Trait Loci) with major effects on the trait
    - Their numbers are limited

Quantitative-Qualitative

- Threshold traits
  - Complex inheritance, polygenic, continuous genetic distribution
  - Discrete phenotypic distribution, Categories
  - Mastitis, inflammation of udders in cows: diseased or not
    - Somatic Cell Count, differently measured in goats and cows
  - Conception rate
    - FSH or LH hormone levels
  - Cats have level of white color ranging from 1-9
    - Controlled by several genes with simple inheritance

Quantitative traits

- Phenotypes have continuous and measurable variation.
- How to measure individual variation in mammals?
  - Weight
  - Wither (the ridge between the shoulder blades of a four-legged mammal) height.
  - BMI= kg/m^2

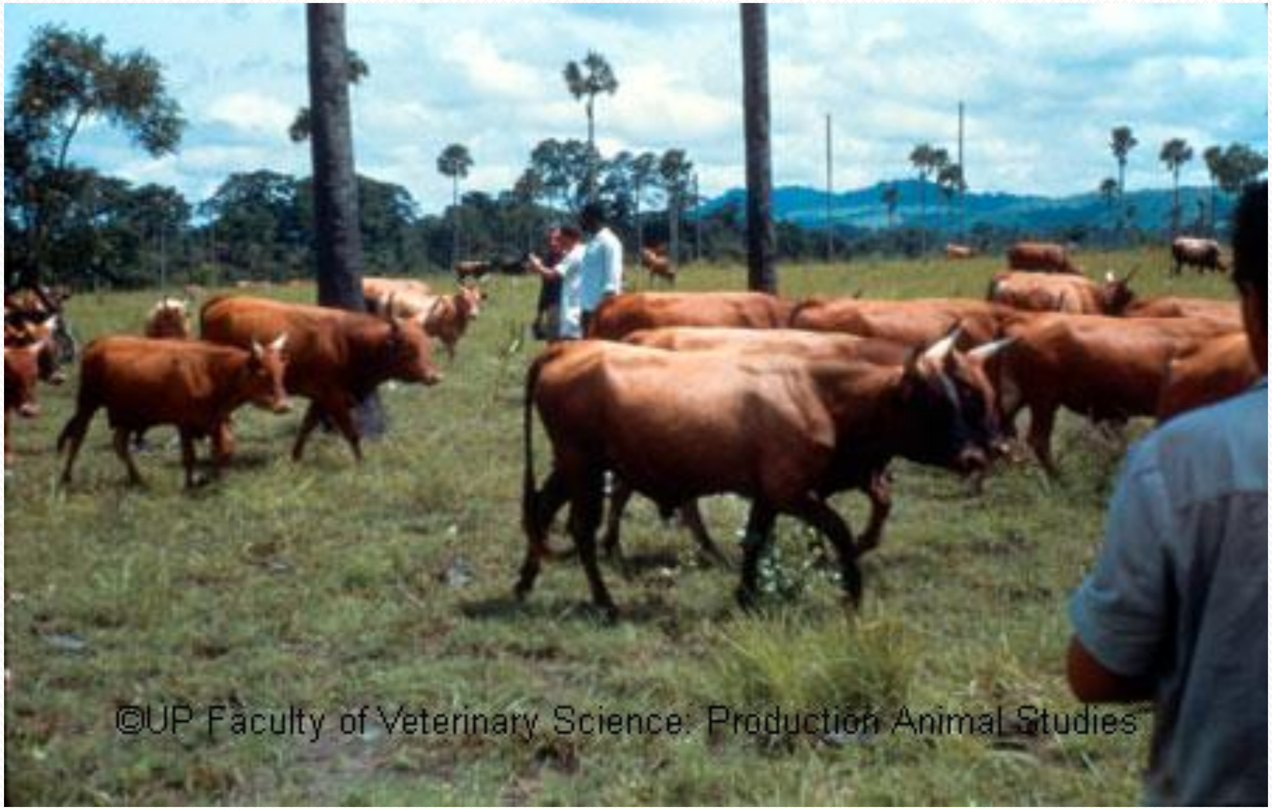
Some Quantitative traits

- **Sheep**
  - Litter size
  - Birth weight
  - 60 day weight
  - Loin Eye area
- **Humans**
  - IQ
  - Blood pressure
  - BMI
- **Dairy Cows**
  - Milk Yield
  - Milk fat (% or gr)
  - Milk protein (% or gr)
  - Calving interval
- **Horses**
  - Racing speed
  - Mature body weight
  - Wither height
- **Plants**
  - Crop Yield
  - Flower diameter

Dwarf Hereford

Dwarfism is a simple inherited qualitative trait. This gene is a major gene in terms of body size. Body size is a quantitative and polygenic trait.

Dwarf Cattle



N'dama cattle (West Africa) are naturally dwarf.

Compared to humans.

Which gene effects are transmitted to the next generation?

- Additive gene effects?
- Dominance?
- Epistasis?

Population Genetics

- Purpose of population genetics is to understand
  - Genetic composition of a population
  - Powers that determines and changes this composition
- Population genetics is the application of Mendel's laws and other genetic principals to a whole population.
- It studies the distributions and changes of allele frequency in a population related to main processes of natural selection, genetic drift, mutation and gene flow (migration).

- Why population genetics is important?

- What does it solve, why is it useful?

Population Genetics

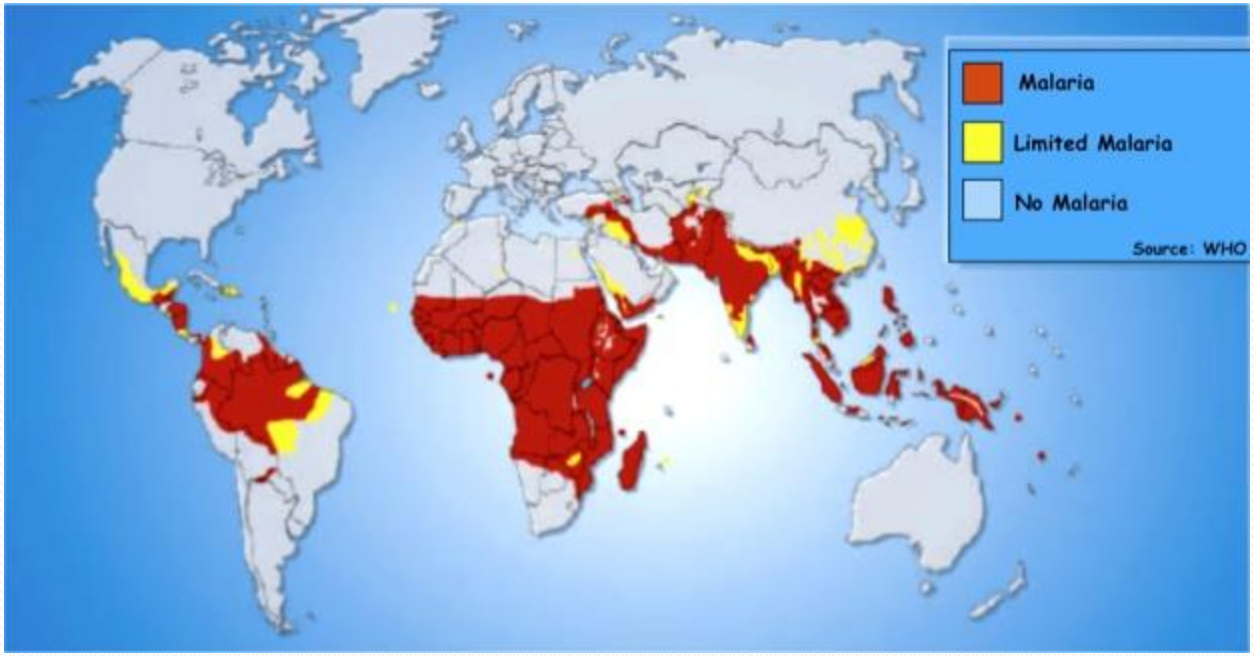
- It studies differences and similarities of individuals.
- When sources of variation is known, transmittance to next generation can be understood.
- Frequency of genes can be calculated to find out what the mean genetic value of a population is.

Gene & Genotype frequency

- Individuals are characterized with genotype,
- Populations are characterized with gene and genotype frequencies for different traits.
- Feather color in chickens can be characterized by genotypes, or feather color allele frequencies or allele combination frequencies.

Variation in some diseases can be explained using population genetics.

- In parts of Africa, this is due to the immunity conferred by the sickle cell gene.

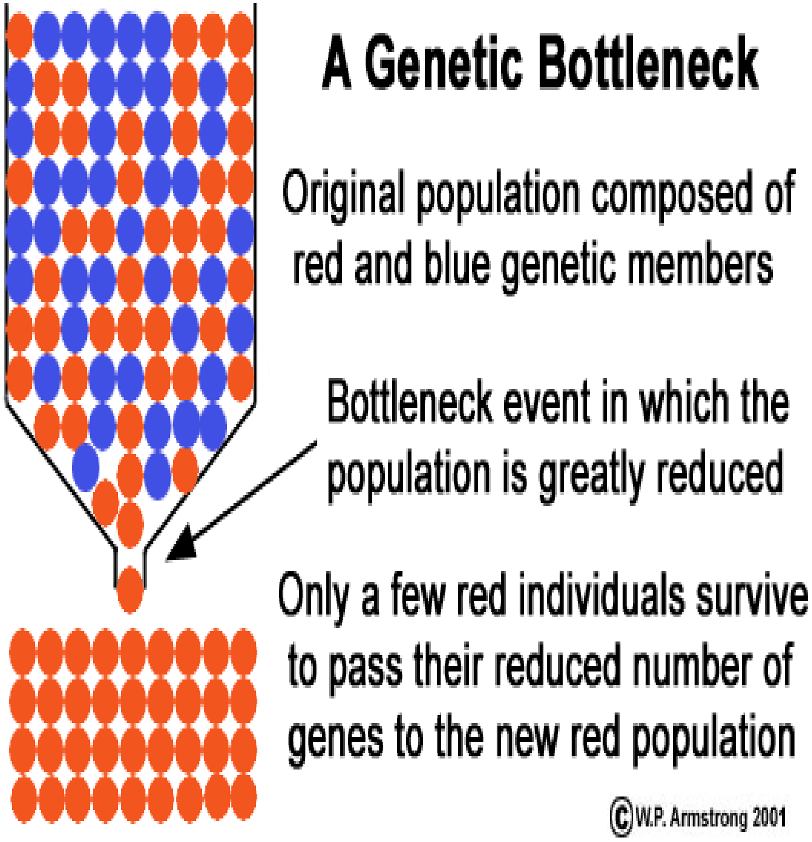


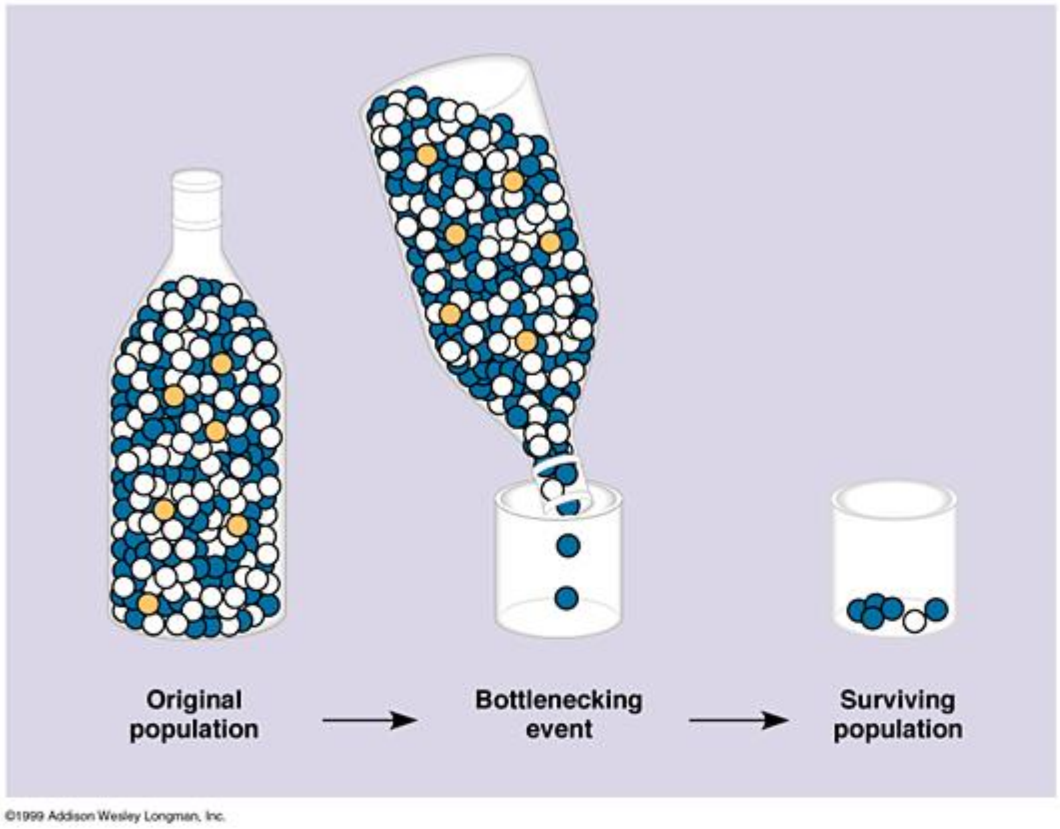
SS normal, dies of malaria (Plasmodia parasite), ss sickle cell anemia, Ss resistant to both malaria and sickle cell anemia.

Sources of changes in genetic variation can be investigated using population genetics. Humans?

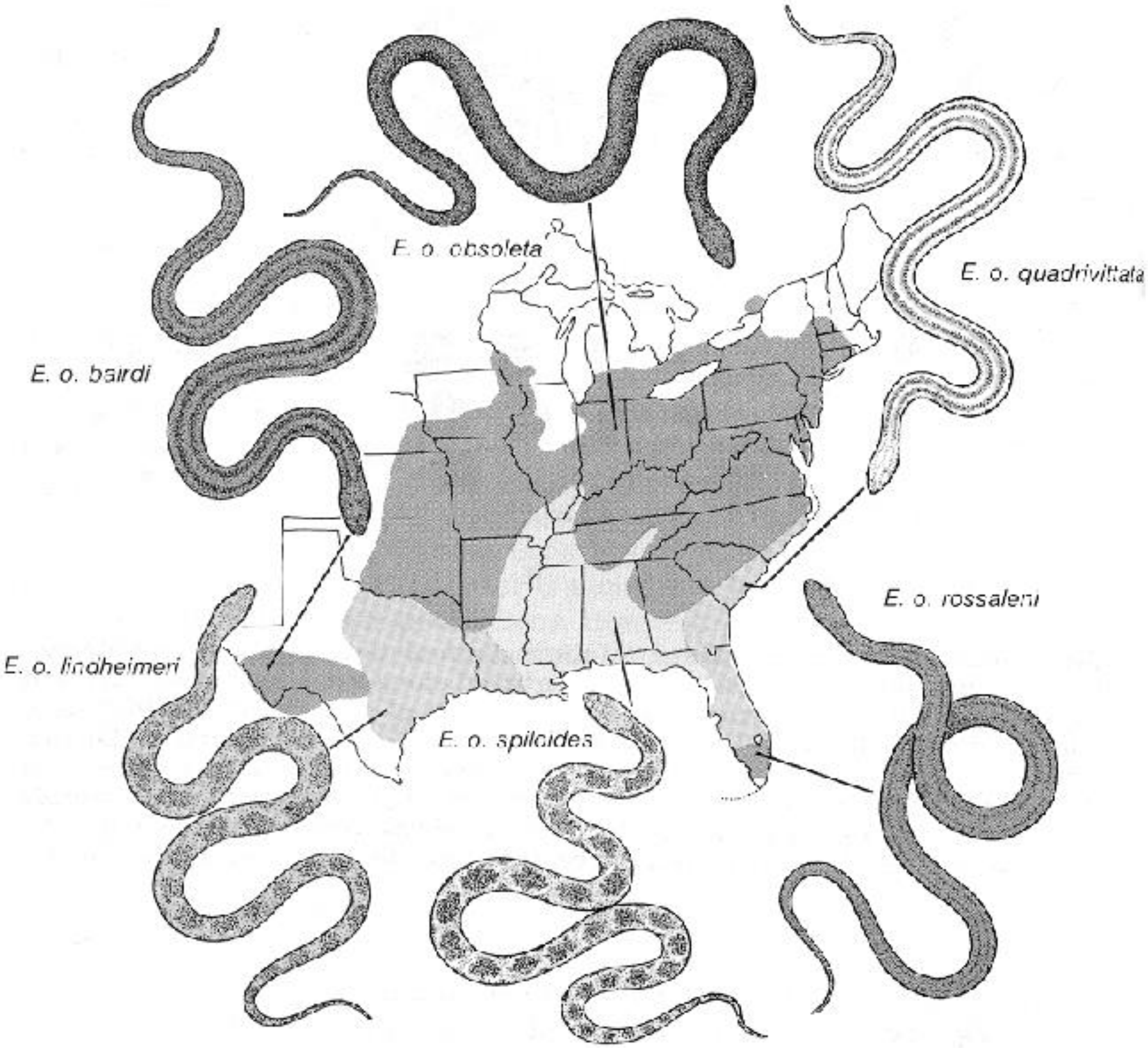
- Cheetahs have the same mouth diameter, focused on gazelles only.
- If gazelles become extinct, cheetahs become extinct as well.
- Humans have large genetic variation.
    - Plenty of migration
    - Me and my brother ☺

Sources of changes in genetic variation can be investigated using population genetics.





Black rat snakes look diverse in different regions



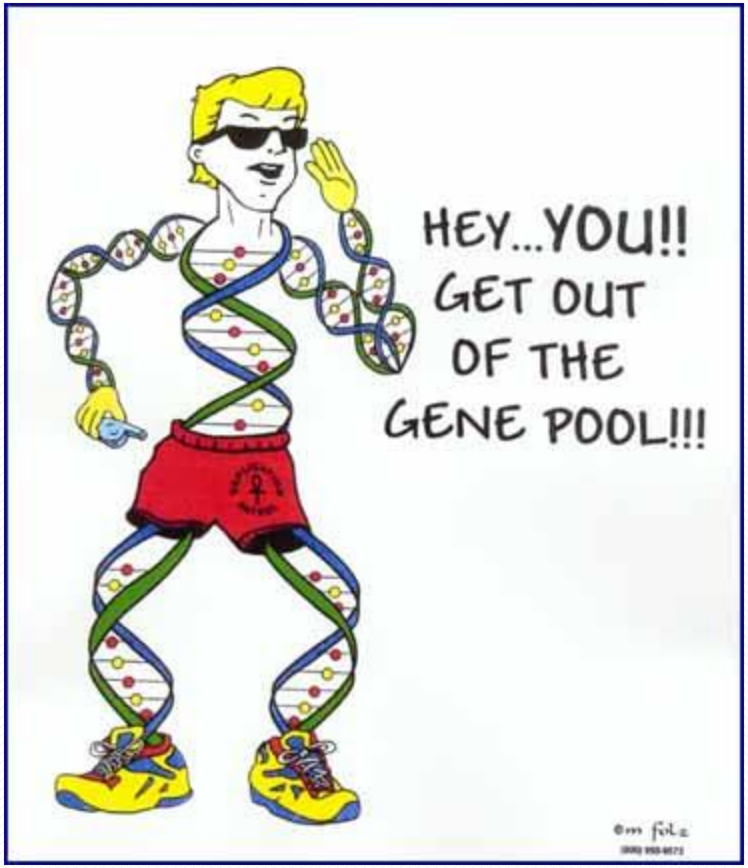
Highly isolated (1500–1900 km apart) populations were strongly divergent, whereas populations more proximal (less-than 465 km apart) although currently isolated, exhibited far less divergence.

K. A. Prior, H. L. Gibbs and P. J. Weatherhead. 1997. Population Genetic Structure in the Black Rat Snake: Implications for Management. Conservation Biology 11, 5, p. 1147–1158.

Mendel Population

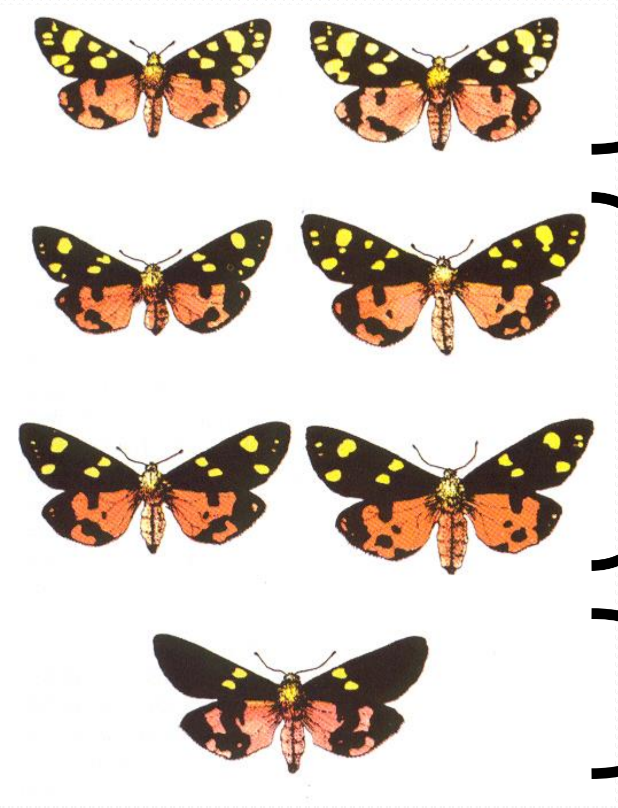
- Basic unit in population genetics is Mendelian population.
- **Mendelian population**, or a local population is an interbreeding group of organisms that share a common gene pool (they carry similar genes).
- **Gene pool** is the total genetic information in an organism population. It is the set of all genes.

One has to understand how a gene pool is formed to understand population genetics.



Describing the gene pool quantitatively

Genotypic frequency



Phenotype	Genotype	Numbers	Frequency
	BB	452	0.909
	Bb	43	0.087
	bb	2	0.004
		Total: 497	

Describing the gene pool quantitatively

Allele frequency

Two alleles, B and b.

$$\boldsymbol{p} = f(\mathrm{B}) = \frac{2\,\mathrm{BB} + 1\,\mathrm{Bb}}{2\,(\mathrm{BB} + \mathrm{Bb} + \mathrm{bb})}$$

$$\frac{\mathrm{BB} + 1/2\,\mathrm{Bb}}{\mathrm{BB} + \mathrm{Bb} + \mathrm{bb}}$$

would also work

$$\boldsymbol{q} = f(\mathrm{b}) = \frac{1\,\mathrm{Bb} + 2\,\mathrm{bb}}{2\,(\mathrm{BB} + \mathrm{Bb} + \mathrm{bb})} = 1 - f(\mathrm{B})$$

$$p + q = 1$$

Describing the gene pool quantitatively

Allele frequency

Genotype	Numbers
BB	452
Bb	43
bb	2
	Total: 497

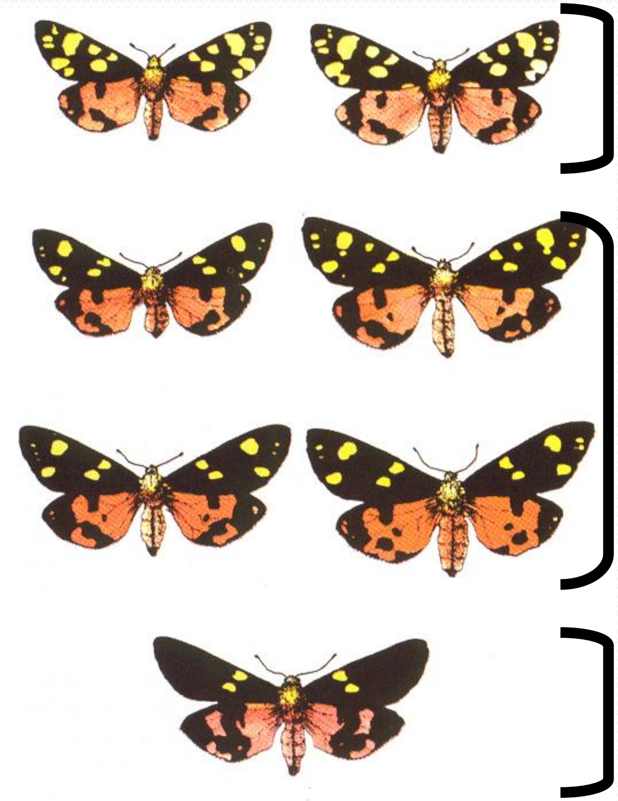
$$\boldsymbol{p} = f(\mathrm{B}) = \frac{2(452) + 1(43)}{2(452 + 43 + 2)} = 0.95$$

$$\boldsymbol{q} = f(\mathrm{b}) = 1 - 0.95 = 0.05$$

Example: $p = f(\mathrm{B}) = 0.95$

$q = f(\mathrm{b}) = 0.05$

Expected genotype frequencies?



genotype

BB $p^2 = 0.95^2 = 0.9025$

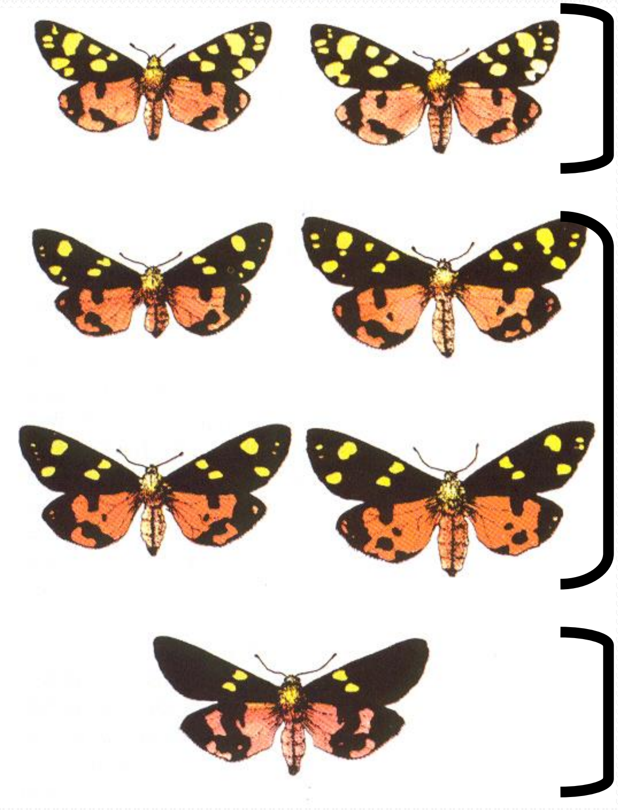
Bb $2pq = 2(0.95)(0.05) = 0.0950$

bb $q^2 = 0.05^2 = 0.0025$

$p = f(\mathrm{B}) = 0.95$

Example: $q = f(\mathrm{b}) = 0.05$

Genotype frequencies?



Genotype	Expected	Observed
BB	0.9025	0.909
Bb	0.0950	0.087
bb	0.0025	0.004

Not an ideal population, thus the difference.
Not infinite population, number of genes is small

Using ***Allele*** frequency to calculate ***genotype*** frequency

$f(\text{A}) = p$
$f(\text{a}) = q$
ise,

$f(\text{AA}) = p \text{ x } p = \boldsymbol{p^2}$
$f(\text{aa}) = q \text{ x } q = \boldsymbol{q^2}$
$f(\text{Aa}) = (p \text{ x } q) = \boldsymbol{pq}$
$f(\text{aA}) = (q \text{ x } p) = \boldsymbol{pq}$

$$f(\text{Aa}) + f(\text{aA}) = (p \text{ x } q) + (q \text{ x } p) = \boldsymbol{2pq}$$

A reminder:

$f(\mathrm{AA}) + f(\mathrm{Aa}) + f(\mathrm{aa}) = 1$

Because total fr. of all genotypes is 1

Thus:

$p^2 + 2pq + q^2 = (p + q)^2 = 1$

Genes with Multiple Alleles

- The Hardy-Weinberg genotype proportions are obtained by expanding the binomial expression to trinomial expression:
- $(p + q + r)^2 = p^2 + q^2 + r^2 + 2pq + 2qr + 2pr$

Genes with Multiple Alleles

Hemoglobin variants in a Nigerian population

AA – 2,017
AS – 783
SS – 4
AC – 173
SC – 14
CC – 11
total – 3,002

$$\boldsymbol{p} = f(\mathrm{A}) = \frac{2(2{,}017) + 783 + 173}{2(3{,}002)} = 0.831$$

$$\boldsymbol{q} = f(\mathrm{S}) = \frac{2(4) + 783 + 14}{2(3{,}002)} = 0.134$$

$$\boldsymbol{r} = f(\mathrm{C}) = 1 - 0.831 - 0.134 = 0.035$$

Calculate r= f(c) to verify

X – linked Genes

$$\boldsymbol{p} = f(X^A) = \frac{2(X^A X^A ♀) + (X^A X^a ♀) + (X^A Y ♂)}{2(♀) + (♂)}$$

$$\boldsymbol{q} = f(X^a) = \frac{(X^A X^a ♀) + 2(X^a X^a ♀) + (X^a Y ♂)}{2(♀) + (♂)}$$

$$\boldsymbol{p + q = 1}$$

Y chromosome does not have many genes, so males cannot be Aa, they are either A or a. We therefore don’t multiply males with 2. Females can be AA or Aa, so it is 2AA and 1Aa.

Population Genetics

- Population genetics starts by characterizing the genetic variation.
- If there are more than one type in a population for a trait, then the population is polymorphic for that trait
- Genetic polymorphism can be seen in every trait.
  - Can you think of a trait/character without genetic polymorphism?
  - # of legs, or # of heads.

Quantification of variation in populations, single locus variation

- Locus-Loci
- Defining genetic structure in one locus is sufficient to define:
  - Which genotypes exist
  - How many of individuals there are from each genotype

Genotype frequencies

- M-N blood groups, Australian aborigines

Blood group	MM	MN	NN	Total
#Individuals	22	216	492	730
freq. of ind.	0.030	0.296	0.674	1.000
Algebraic illustration	P	H	Q	P+H+Q=1

http://onlinelibrary.wiley.com/doi/10.1002/ajpa.1330120409/abstract

Variation

- Genotype frequencies describe a population sufficiently, but does not describe a reproducing population with enough information
- That's because genes are transmitted to the next generation, not genotypes.

Gene Frequencies (transmitted to next generation)

- M-N blood groups, Australian aborigines

Blood group		MM	MN	NN	Total
#Individuals		22	216	492	730
# Genes	M	44	216	0	260
	N	0	216	984	1200

Gene frequencies

- Frequency of M allele
- f(M)=#of M genes/# of total genes
  - =260/1460=0.178=p
- f(N)=)=#of N genes/# of total genes
  - =1200/1460=0.822=q

Genotype frequencies

- Gene frequencies can be calculated using genotype frequencies.
- $p=f(M)=P+H/2=0.030+0.296/2=0.03+0.148=0.178$
  - given as freq. of ind.=freq. of genotypes in this case because heterozygous genotypes can be observed (codominance)
  - Similar to snapdragon of red, pink and white (incomplete dom.)
    - Normally in a complete dominance you would only see two types such as black cattle and red cattle
- $f(N)=Q+H/2=0674+0.296/2=0822$

Factors affecting genetics of a population

- Factors causing short term changes in allele fr.
  - Migration
  - Natural/artificial selection
  - Population structure (size and mating systems)
    - Genetic drift, subpopulations
- Factors causing long term changes in allele fr.
  - Mutation
  - Migration (large genetic flow)
  - Natural/artificial selection (tight selection)
  - Population structure (size and mating systems)
    - Genetic drift (long time drift)

Factors & genetic variance

- Factors increasing genetic variance
  - Mutation
  - Migration
- Factors decreasing genetic variance
  - Population structure
    - Genetic drift
    - Effective population size is important
      - How many females a male is mating with?
  - Inbreeding may decrease variance in a small population if it is used along with selection.
  - Natural/artificial selection (tight selection) may decrease s^2 if heritability and the accuracy is high (BLUP mates relatives)

Hardy-Weinberg equilibrium

- Purpose of population genetics
  - Investigate how these factors affect genetic variation separately and together (interaction)
  - Their importance in changes of gene frequency.
- Relationships between allele frequencies and genotype frequencies are explained using Hardy-Weinberg.
- Using the allele frequencies (p, q), we can get genotype frequencies in an ideal population.
- The Hardy–Weinberg equilibrium, also known as the Hardy–Weinberg principle, model, theorem, or law, states that allele and genotype frequencies in a population will remain constant from generation to generation.

Idealized Population

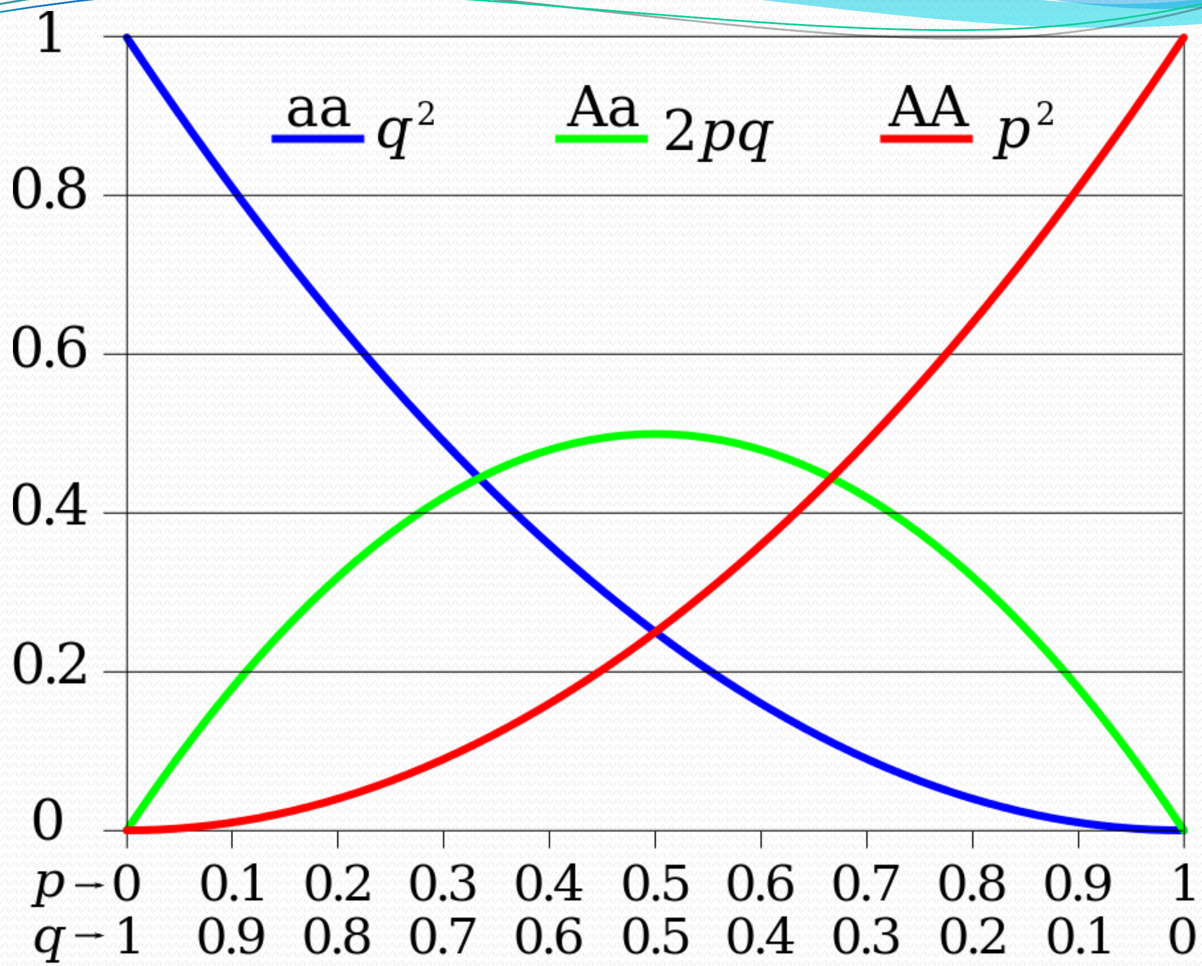
- First off, we start by defining an idealized population where none of these factors play a role,
  - use a number of simplifying assumptions.
- Then each of these factors are included one by one.
- The first idealized population model was created in 1908 y. by G. H. Hardy and Wilhelm Weinberg:
- Allele and genotype frequencies in a population will remain constant from generation to generation.
  - Can estimate genotype frequencies using parent generation gene frequencies.

Ideal Population

- The organism is diploid
- Reproduction is with sex, gene frequency is the same in males and females (no X linked genes)
- Generations are not overlapping (Parent and offspring generation does not mate)
- Mating is random (only for the trait considered)
- Population size is infinite (small size has drift)
- No migration
- No mutation
- No selection
  - NO SUCH NATURAL POPULATION EXISTS ☺

Ideal Population is

1. Stable in terms of gene and genotype frequencies.
   - Genetic features do not change between generations. Mendel's segregation law and random mating keeps the variance intact when other factors are not in play.
2. Gene frequencies in the parental generation determines the genotype frequencies in the offspring generation.
   - If parental gene fr. are p and q,
   - Offspring generation genotype fr. are $p^2 + 2pq + q^2$



Hardy–Weinberg proportions for two alleles: the horizontal axis shows the two allele frequencies p and q and the vertical axis shows the expected genotype frequencies. Each line shows one of the three possible genotypes.

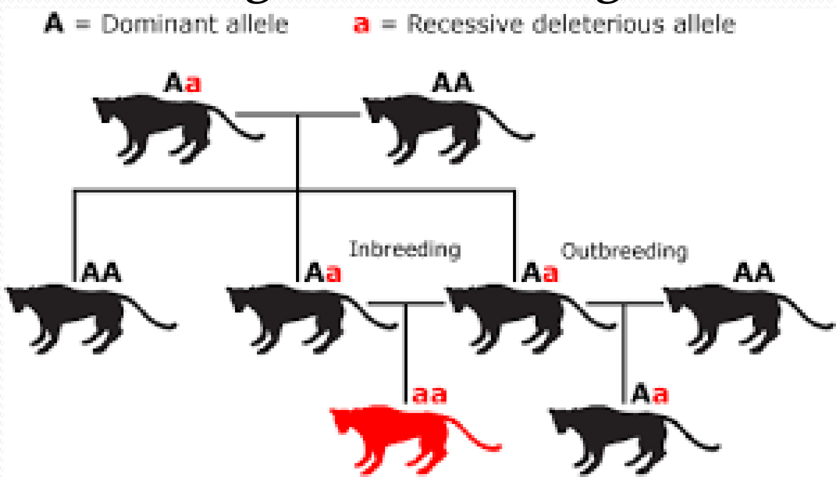
Exceptions to the Hardy-Weinberg Principle

- Nonrandom mating (including selection)
- Unequal survival
- Population subdivision
  - causes genetic drift because of small size
- Migration

will disrupt Hardy-Weinberg equilibrium

Nonrandom Mating

- Nonrandom mating includes
  - Inbreeding
    - Mating between <u>genetically</u> related individuals
  - Assortative mating
    - Mating between individuals with similar <u>phenotypes</u>
- Both inbreeding and assortative mating <u>reduce the frequency of heterozygotes</u> and increase the frequency of homozygotes compared to the Hardy-Weinberg genotype frequencies.
- The effects of inbreeding can be quantified using the inbreeding coefficient, ***F***.
- If F=1 Self-fertilization
  - not selfie ☺



Unequal Survival

- If zygotes produced by random mating have different survival rates....

Heterozygotes ≠ Homozygotes

- Which one of them do you expect to have higher survival?
  - Of course heterozygotes
- A sample of 200 adults yielded the following data:

Genotype	Observed Count		Expected Count
A_1A_1	26		46.1
A_1A_2	140		99.8
A_2A_2	34	higher than expected	54.1

Population Subdivision

- There are no mating restrictions at all
  - It is **panmictic**
- Panmixis implies that any member of the population is able to mate with any other member.
- In nature, populations may be subdivided due to geographical or ecological barriers that may be correlated with genetic differences.
- Because they are small, subdivisions may be under the influence of genetic drift.

Migration (gene flow)

- The introduction of genes by recent migrations can alter **allele and genotype frequencies** within a population and disrupt Hardy-Weinberg equilibrium.

Genetic drift=chance

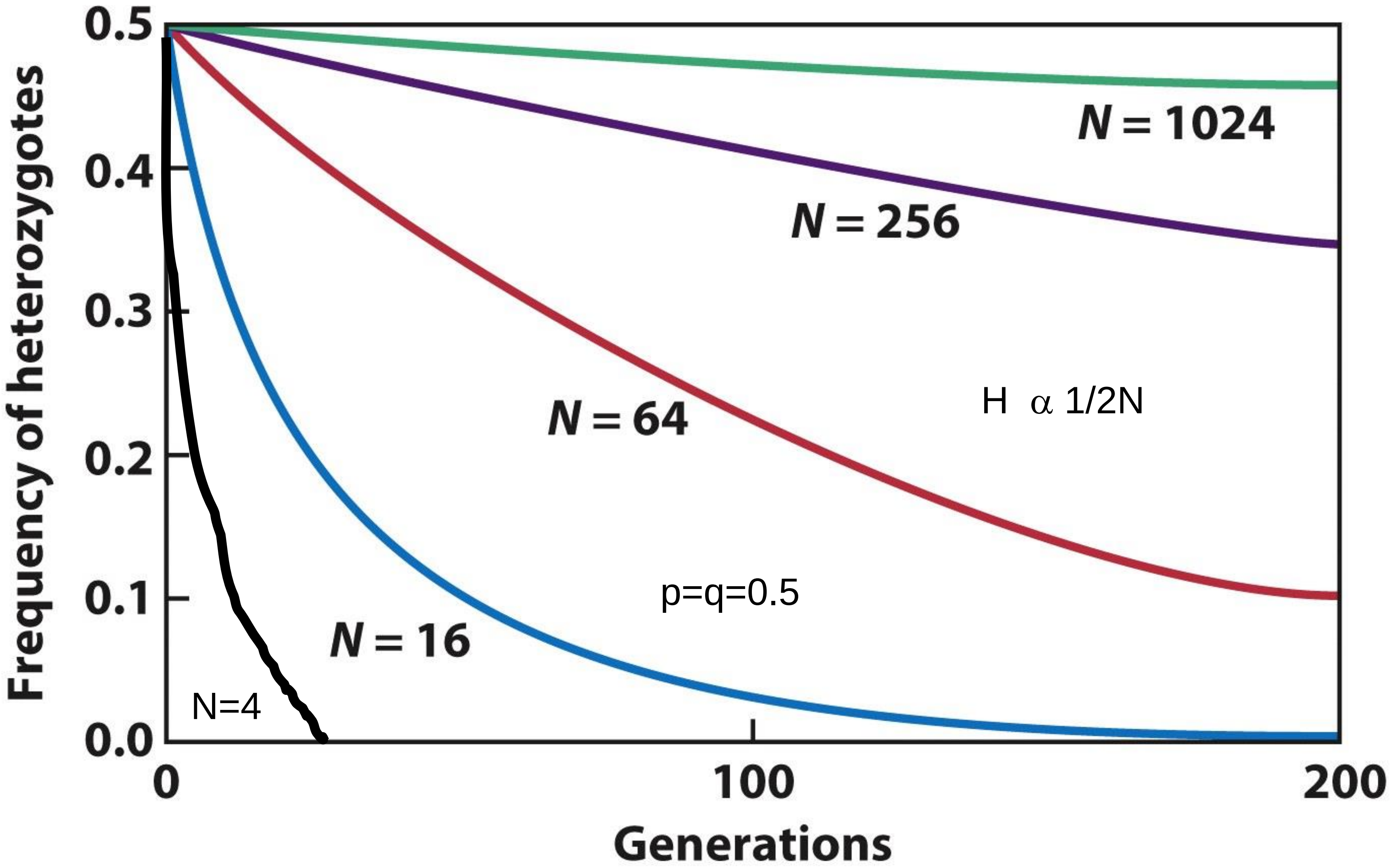
- Genetic drift is the random change of allele frequencies in populations.
- It is due to uncertainties in Mendelian segregation.
- Allele frequencies change unpredictably in populations because of uncertainties during reproduction.
  - You don't know which genes will go to the progeny.

Factors Contributing to Random Genetic Drift

- There is always **uncertainty** as to which allele a given offspring will receive.
- There is **random variation** in the <u>number</u> of offspring that a parent produces.

The Effects of Population Size

- In large populations, the effect of genetic drift is minimal.
- In small populations, genetic drift may be the primary change force.
- The effect of population size is determined by monitoring the frequency of heterozygotes, or the **heterozygosity** of a population over time.
  - Effective population size (Ne), is related to number of mating males and females



Frequency of Heterozygotes with different sample sizes

Ideal population

- There are 3 steps for the Hardy-Weinberg equilibrium of gene and genotype frequencies
  1. From parents to the gametes they produce
  2. From random gametes to zygote genotype
  3. From zygote genotype frequency to the offspring gene frequency.

1. From parents to the gametes they produce

- Suppose we have A locus and two alleles (A_1, A_2), with the frequencies p and q.
- Three genotypes, A_1A_1, A_1A_2 and A_2A_2
- with frequencies P, H and Q.
- Gamete frequencies=gene frequencies,
  - $f(A_1) = p = P + H/2$
  - $f(A_2) = q = Q + H/2$

2. From random gametes to zygote genotypes

Zygote genotype frequencies

- $A_1A_1 = p^2$
- $A_1A_2 = 2pq$
- $A_2A_2 = q^2$

		Male Gamete A_1 (p)	Male Gamete A_2 (q)
Female Gamete	A_1 (p)	A_1A_1 (p^2)	A_1A_2 (pq)
	A_2 (q)	A_1A_2 (pq)	A_2A_2 (q^2)

3. From zygote genotype frequency to the offspring gene frequency

- $f(A_1) = P + H/2$

$$
\begin{aligned}
&= p^2 + 2pq/2 \\
&= p^2 + pq \\
&= p(p + q) \qquad\qquad p + q = 1 \\
&= p
\end{aligned}
$$

- This shows that gene frequencies in the offspring generation is the same as the gene frequencies in the parental generation
  - If there is no mutation, migration, selection, drift etc.
  - Meaning if the population is in HW equilibrium

An example: Cystic Fibrosis

- Cystic Fibrosis (CF) is a genetic and chronic disease
- One of the most common disease among Caucasians.
- A mutation in the CF gene causes thickness and stickiness in the mucus.
- Causes digestion and respiration problems.
- in 1955, a child with CF was not expected to live much beyond the age of five, the median was 6 months. Today, dramatic advances in CF research and treatment have helped to extend the median age of survival for a person with CF to median age of 37.

Cystic Fibrosis

- Frequency of the disease is 0.0006 (1 in 1667 ind.).
- CF is an autosomal recessive disorder caused by the presence of mutations in both copies of the gene.
- If the human population is assumed to be in Hardy-Weinberg equilibrium,
- $0.0006=q^2$ and therefore $q=0.0245$
- Thus, frequency of heterozygote carriers:
- $2pq=2(1-0.0245)(0.0245) = 0.0478$

Cystic Fibrosis

- Every 1 in 1667 people has this genetic disease.
- Every 1 in 21 people is heterozygous carrier.
- Majority of the recessive alleles are in phenotypically normal people.
- Every human is heterozygous carrier for 7-8 deleterious recessive gene.
- Importance for inbreeding?
  - Individuals with similar genes (relatives) have increased probability for offspring with homozygous bad genotype

Predicting Allele Frequencies from Genotype Frequencies

- The incidence of the **autosomal recessive** disorder phenylketonuria (PKU) is about 0.0001 (1 in 10000).
  - This number is 0.00025 (1 in 4000) in Turkey
- The incidence of PKU represents the frequency of mutant homozygotes in the population.

- Homozygous mutant individuals should occur with a frequency equal to the square of the mutant allele frequency, q.
- $q^2 = 0.0001$
- Taking the square root, $q = 0.01$
- Because $p + q = 1$, we know that $p = 0.99$.
- $p = 0.99$ and $q = 0.01$
- The frequency of heterozygous carriers is $2pq = 2(0.99)(0.01) = 0.0198$.
- **For Turkey**: $\sqrt{0.00025}=0.0158$
- 2pq=2(1-0.0158)(0.0158)=2*0.9841*0.0158=0.03
- 3 % of Turkish people are heterozygous carriers

Problem

- Given the genotype frequencies, find whether the population is under the HW equilibrium.

- MN blood groups genotype frequencies:

Genotype	**MM**	**MN**	**NN**	**Total**
Observed	22	216	492	730

Problem, step 1

- Calculate the observed gene frequencies
- p=frequency (M)

$$= \frac{(\#\ \text{MM homozygous}) + (\#\ 1/2\ \text{MN heterozygous})}{\text{Total}}$$

$$= \frac{(22) + (\frac{1}{2} \times 216)}{730} = 0.178$$

Problem, step 1

- Calculate the observed gene frequencies
- q= frequency (N)

$$= \frac{(\#\ \text{NN homozygous}) + (\#\ 1/2\ \text{MN heterozygous})}{\text{Total}}$$

$$= \frac{(492) + (\frac{1}{2} \times 216)}{730} = 0.822$$

or, q=1-p=1-0.178=0.822

Problem, step 2

- Using the observed gene frequencies, calculate the genotype frequencies expected under Hardy-Weinberg

Genotype	MM	MN	NN	Total
Expected genotype frequencies	p^2	$2pq$	q^2	1
	$=(0.178)^2$	$=2(0.178)*(0.822)$	$=(0.822)^2$	
	=0.0317	=0.2926	=0.6757	1

Problem, step 3

- Calculate expected genotype count by multiplying expected genotype frequencies * sample size.

Genotype	MM	MN	NN	Total
Expected genotype count	p^2*total	2pq*total	q^2*total	730
	=0.0317*730	=0.2926*730	=0.6757*730	
	=23.141	=213.598	=493.261	730

Problem, step 4

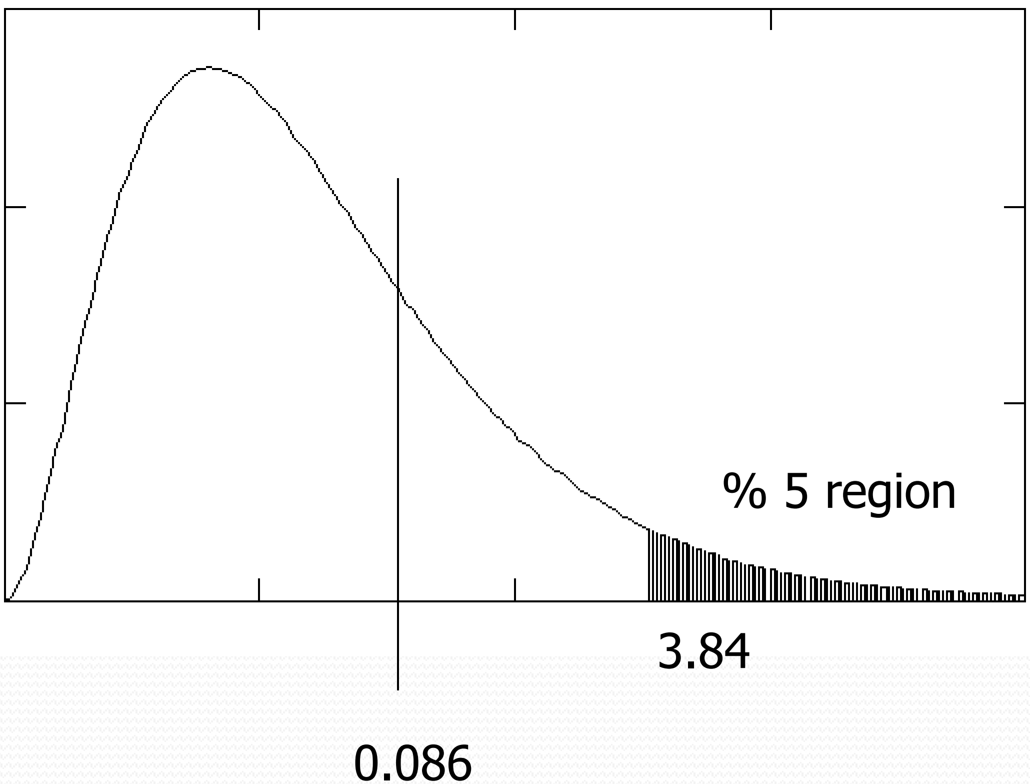
- Compare observed genotype count and expected genotype count using χ^2

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

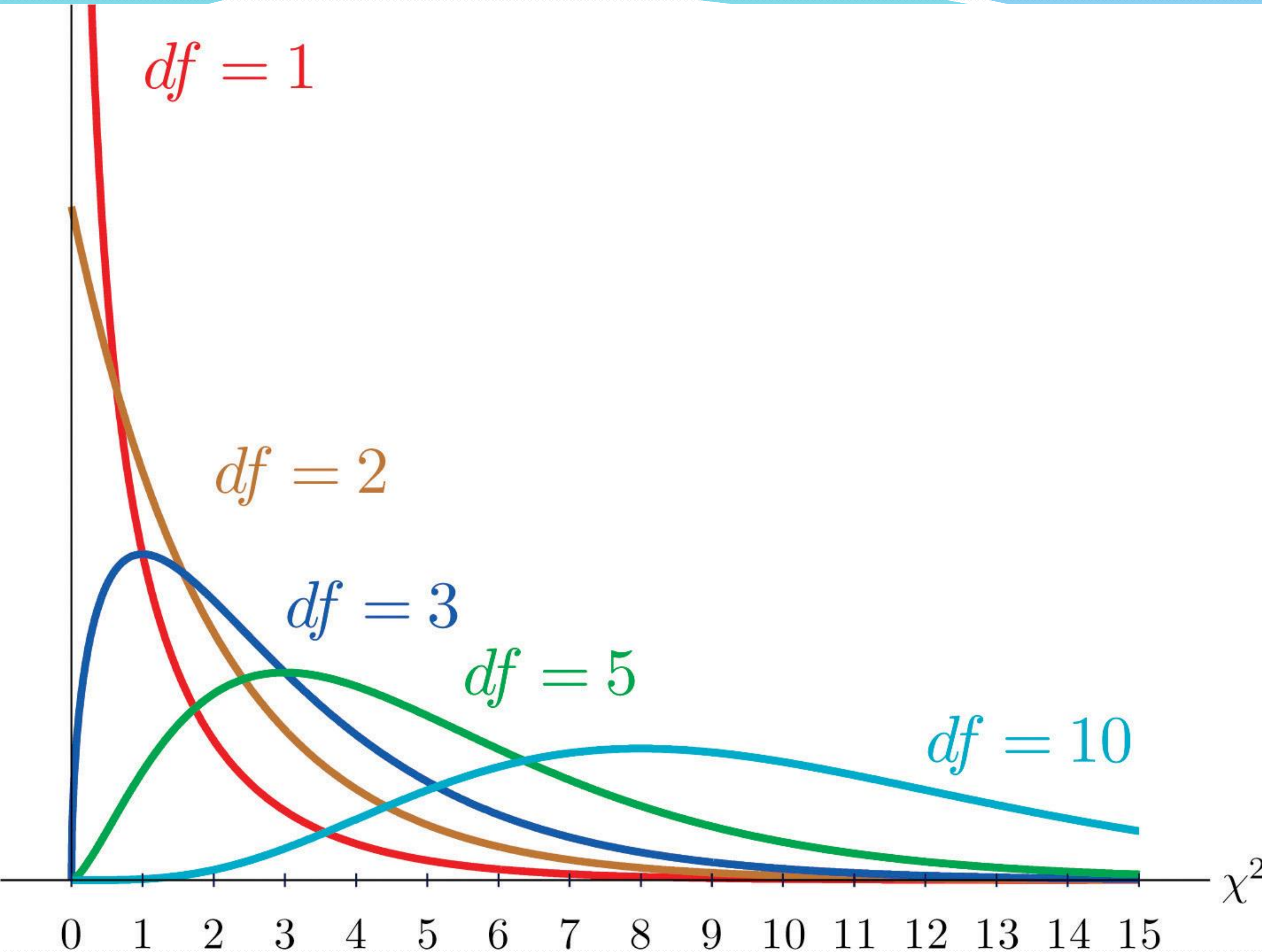
$$\chi^2 = \frac{(22-23.141)^2}{23.141} + \frac{(216-213.598)^2}{213.598} + \frac{(492-493.261)^2}{493.261}$$

Result

- Chi square value= 0.086



Difference is due to chance, H_0 hypothesis is not rejected.



1 degrees of freedom chi square is normally as seen on left

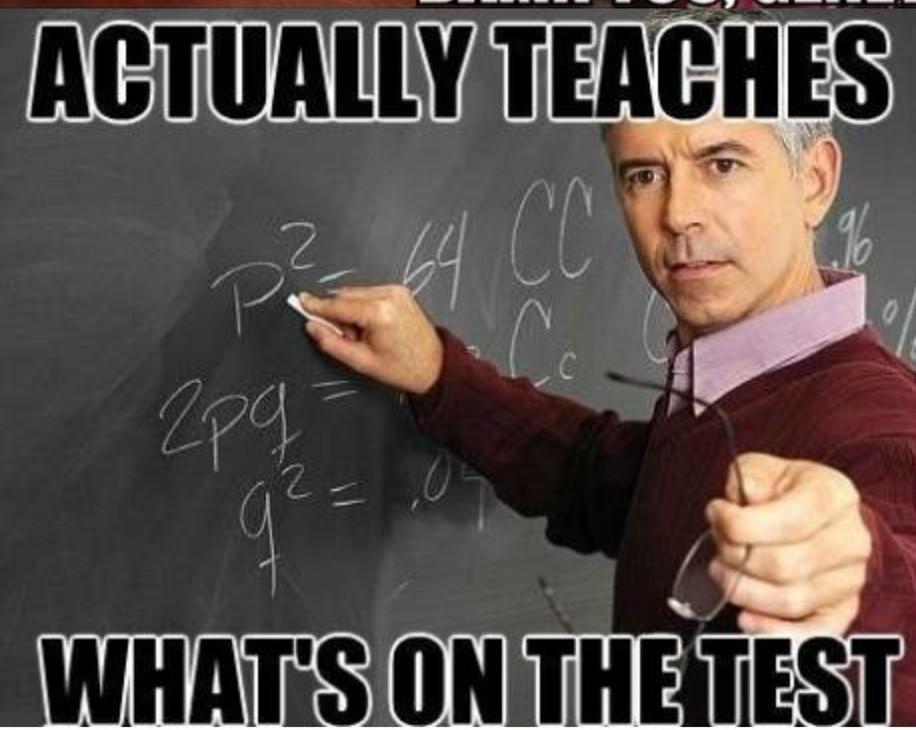
Result

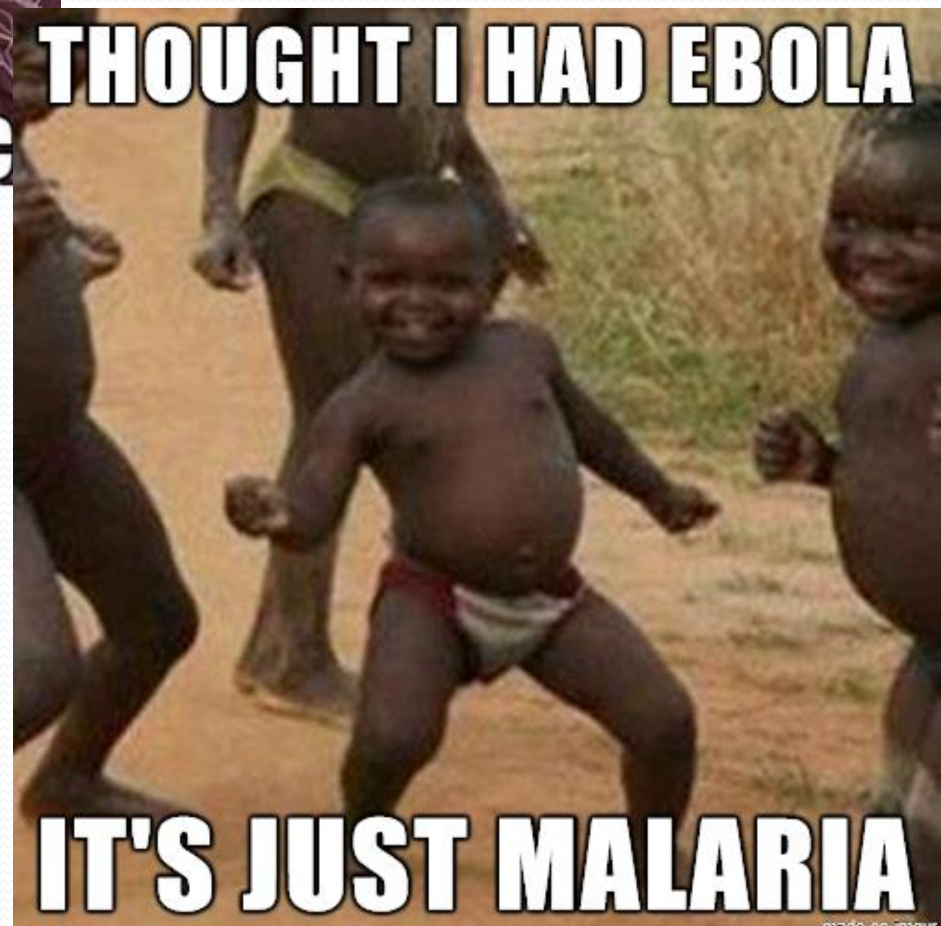
- Expected and observed values are very close (22-23, 216-213, 492-493).
- Calculated chi square value, using the 0.05 probability level as our critical value and 1 df, is compared with the table chi square value,
- Table χ^2 (1df)= 3.84
- Degrees of freedom=# of classes-1(total)-1(estimated every parameter)
- In our example df=3(class)-1(total)-1(gene frequency)=1

Conclusion

- Calculated chi square value 0.086, is lower than the table value of 3.84, which means that we failed to reject H_0 hypothesis (accept it) and this conclusion is reached:
- This population is under Hardy-Weinberg equilibrium for the MN blood group locus
  - or we failed to prove that it is not.


DAMN YOU, GENETIC


ACTUALLY TEACHES
WHAT'S ON THE TEST


THOUGHT I HAD EBOLA
IT'S JUST MALARIA
made on imgur