

# The Hardy-Weinberg principle

- Description of variation at one gene
  - Genotype/allele frequencies
- The Hardy-Weinberg principle
  - Autosomal gene with 2 codominant alleles
  - Statistical test
  - Sex-linked genes
  - Autosomal gene with a dominant allele
  - Graphical test for multiple diallelic genes

# Reminder: Locus, gene, allele

## Gene

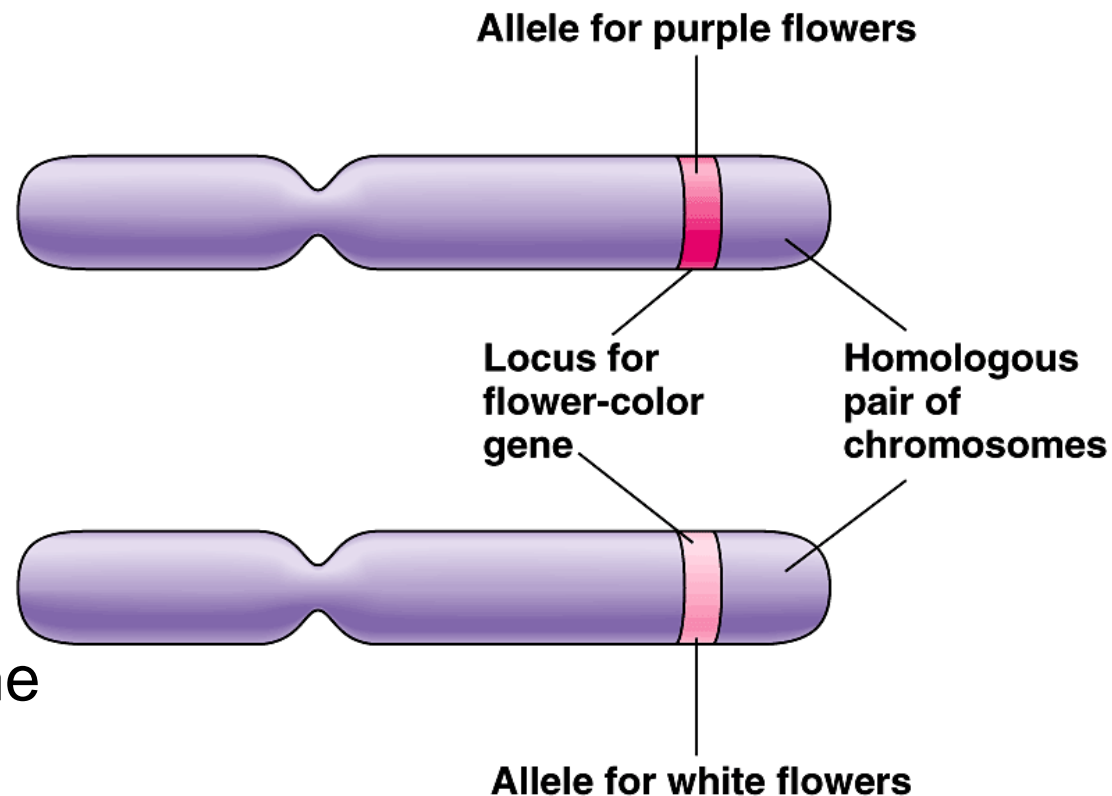
Codes for a specific trait  
(flower color)

## Allele

Specific form of gene  
(Purple, white)

## Locus

Genes' position in genome  
(plural loci)



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

# Reminder: Dominant, recessive, codominant, incomplete dominant

## Diploid organism

**Genotype**

$A_1A_1$

$A_1A_2$

$A_2A_2$

**Phenotype**

Red

Red

White



$A_1$  dominant,  $A_2$  recessive

# Reminder: Dominant, recessive, codominant, incomplete dominant

## Diploid organism

**Genotype**

$A_1A_1$

$A_1A_2$

$A_2A_2$

**Phenotype**

Red

Red and white

White



$A_1$  and  $A_2$  are codominant

# Reminder: Dominant, recessive, codominant, incomplete dominant

## Diploid organism

**Genotype**

$A_1A_1$

$A_1A_2$

$A_2A_2$

**Phenotype**

Red

Pink

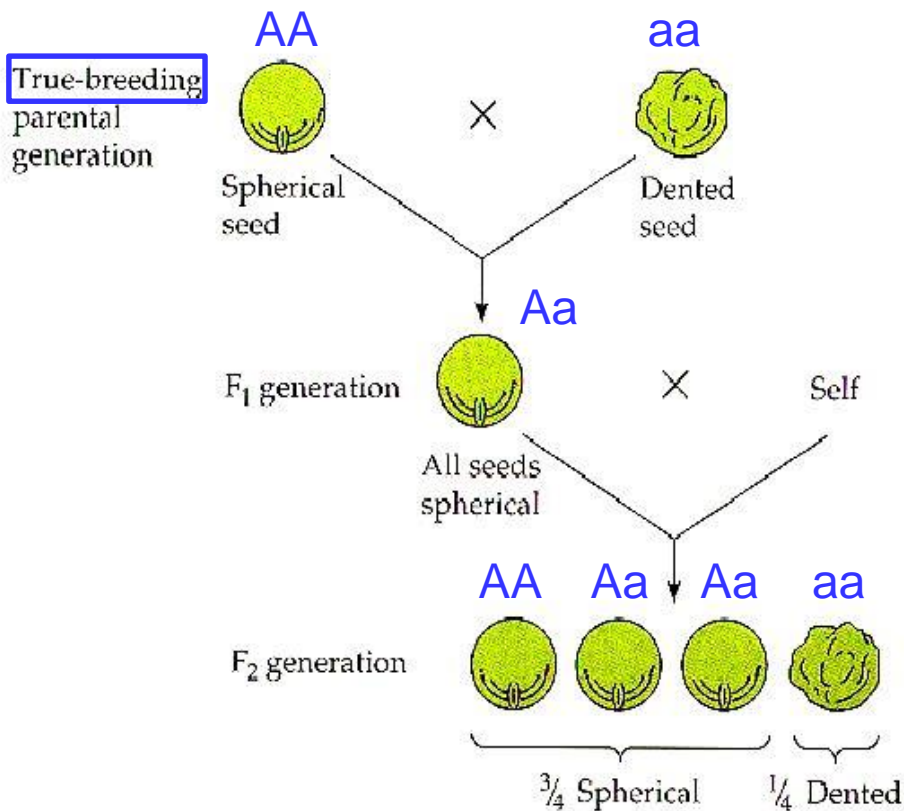
White



Incomplete dominance

# Analysis of genetic variation of a gene in an experimental population

## Mendel's laws in 1866



## Assumptions:

- 1) diploid organism
- 2) autosomal gene
- 3) 2 alleles
- 4) dominance among alleles
- 5) 3 + 4  $\Rightarrow$  2 phenotypes



# Analysis of genetic variation of a gene in a natural population

In a plant species the color of the flowers is determined by two alleles, one is dominant and the other recessive.

In a population of this species, 75% of the individuals carry red flowers and 25% white flowers.

What is the frequency of the allele coding for white flowers?

# Analysis of genetic variation at a locus in a natural population

SNP (**S**ingle **N**ucleotide **P**olymorphism)

Allele 1 ...GATCCGTACTG**A**TGGGATGG...

Allele 2 ...GATCCGTACTG**G**TGGGATGG...



# Analysis of genetic variation at a locus in a natural population

SNP (**S**ingle **N**ucleotide **P**olymorphism)

Example from

Exome Aggregation Consortium (ExAC) database  
( $n > 60,000$ )

Exome: part of the genome composed of exons  
(1% of genome,  $\approx 30,000,000$  bp)

SNP rs509360 located on Chromosome 11: 61781087

Population: East Asians

# Genotype and allele frequencies

Genotype	AA	AG	GG	Sum
Number	968	1914	1424	4316
Frequency	0.2243	0.4434	0.3323	1
Frequency	$P_{AA}$	$P_{AG}$	$P_{GG}$	1

## Allele frequencies

$$p = (2 \times 968 + 1914) / (2 \times 4316) = 0.446$$

$$q = (2 \times 1424 + 1914) / (2 \times 4316) = 0.554$$

$$p = P_{AA} + P_{AG} / 2$$

$$q = P_{GG} + P_{AG} / 2$$

$$p + q = 1$$

# The Hardy-Weinberg principle

What is the genotype distribution at an autosomal locus in a natural population?

The principle was derived in 1908

G.H. Hardy, English mathematician  
(1877–1947)



W. Weinberg, German physician  
(1862–1937)



# The Hardy-Weinberg principle

- Diploid organism, animal
- Autosomal locus, 2 alleles
- Identical genotype distribution in males and females
- Non-overlapping generations
- Infinite population
- No mutation
- No migration
- No selection
- Random mating (panmictic)
- Mendelian segregation

Genotype	$A_1A_1$	$A_1A_2$	$A_2A_2$
Adults	$P_{11}$	$P_{12}$	$P_{22}$
Offspring	??	??	??

# The Hardy-Weinberg principle

		Sperm	
		$\text{fr}(A_1) = p$	$\text{fr}(A_2) = q$
Eggs	$\text{fr}(A_1) = p$	$\text{fr}(A_1A_1) = p^2$	$\text{fr}(A_1A_2) = pq$
	$\text{fr}(A_2) = q$	$\text{fr}(A_2A_1) = qp$	$\text{fr}(A_2A_2) = q^2$
		Zygotes	

# The Hardy-Weinberg principle

Genotype	$A_1A_1$	$A_1A_2$	$A_2A_2$
Adults	$P_{11}$	$P_{12}$	$P_{22}$
Offspring	$p^2$	$2pq$	$q^2$

Allele frequency among the offspring

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

Allele frequency is constant

# Hardy-Weinberg equilibrium

Independent of the genotype frequencies in  
a population,

after one generation with random mating  
and Mendelian segregation

the population is in Hardy-Weinberg  
equilibrium with the genotype distribution

$A_1A_1$

$A_1A_2$

$A_2A_2$

$p^2$

$2pq$

$q^2$

and remains there

# Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium is a **mixed** equilibrium.



# Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium is **neutral** with respect to allele frequencies.

## Equilibrium types

Stable



Unstable



Neutral



# Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium is **neutral** with respect to allele frequencies.

If the allele frequencies change to  $p'$  and  $q'$ , they will remain there in the future.

$$A_1A_1$$

$$p^2$$

$$A_1A_2$$

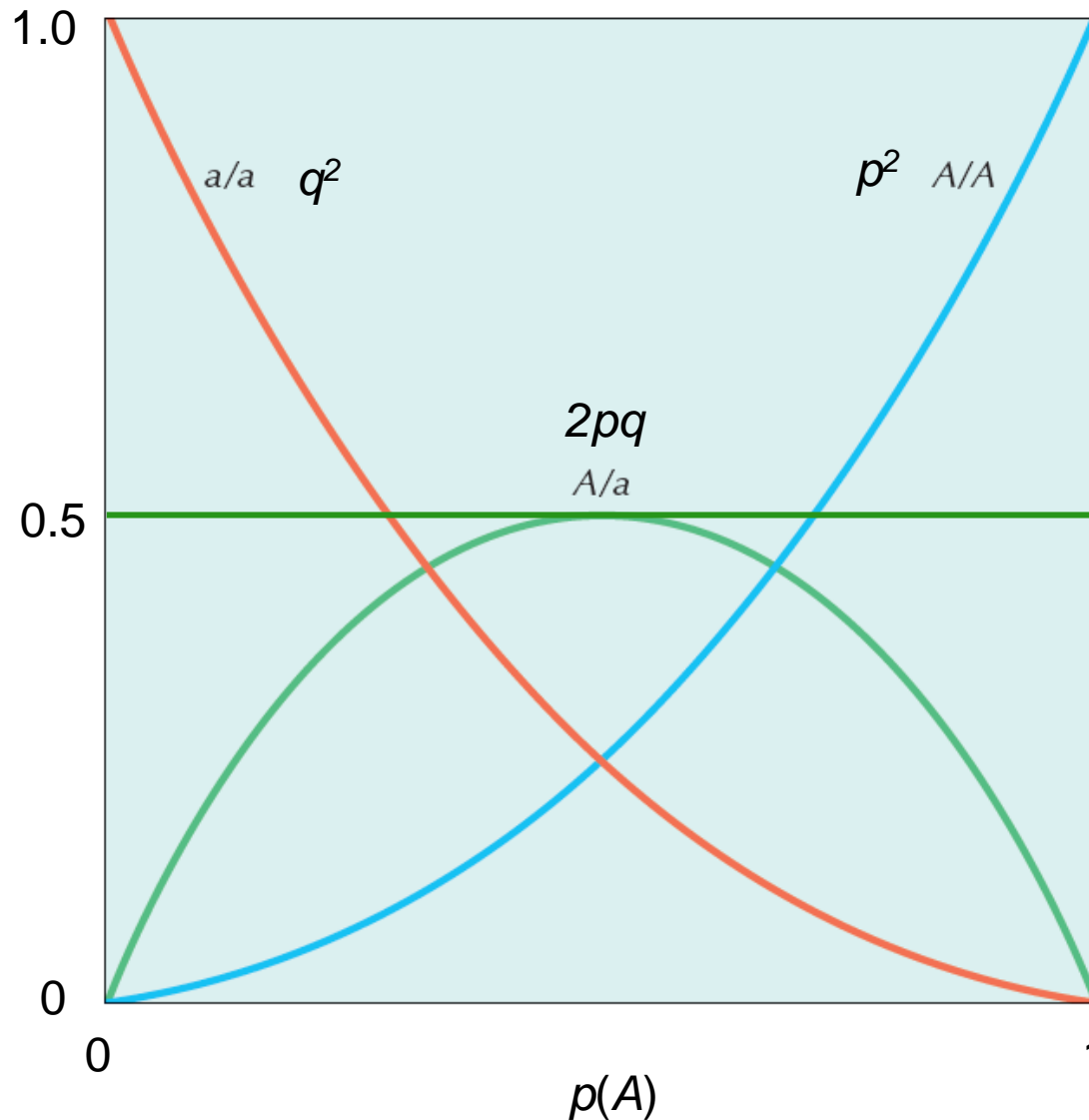
$$2p'q'$$

$$A_2A_2$$

$$q'^2$$

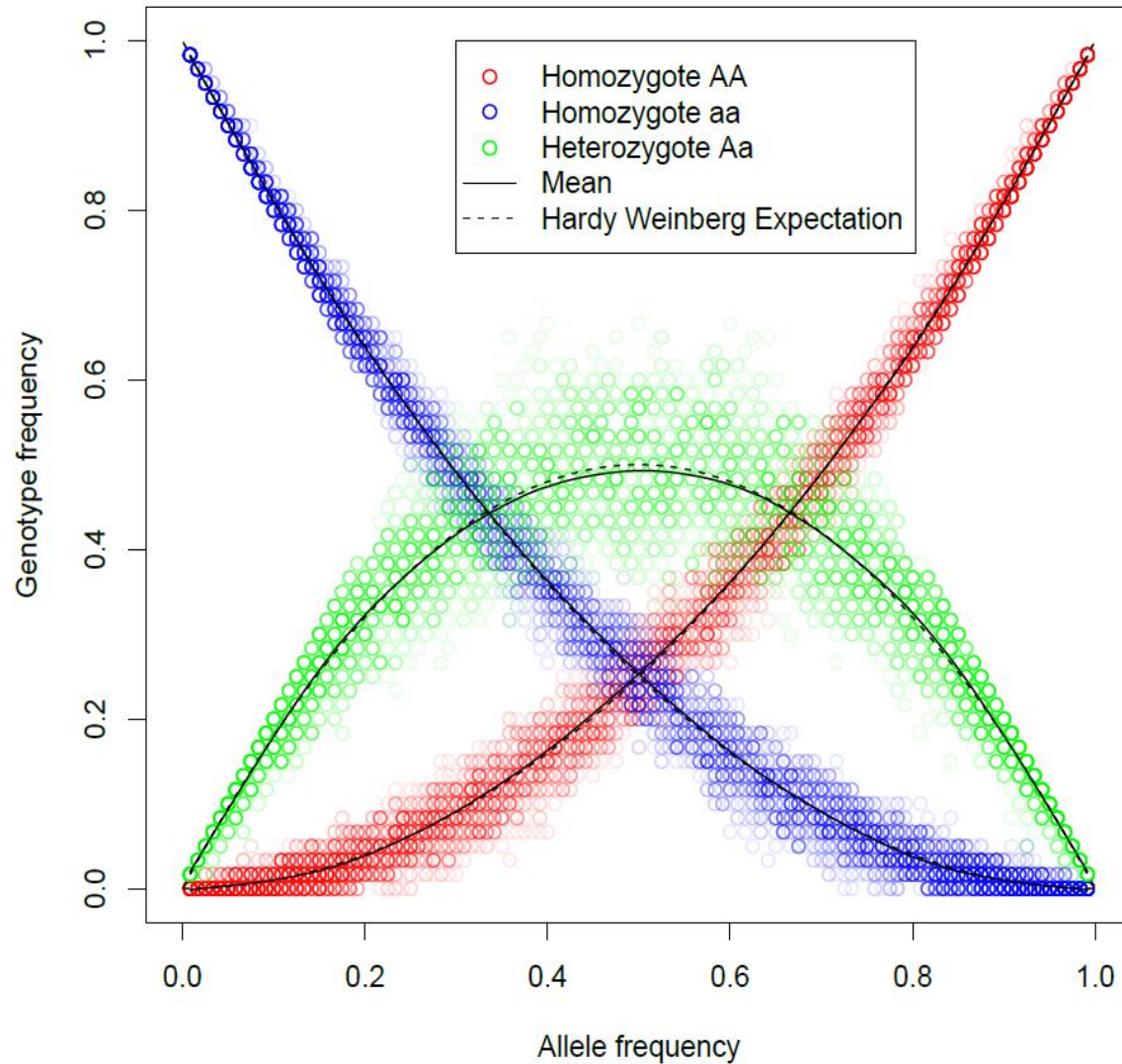
The genotype frequencies for given allele frequencies are **stable**.

# Hardy-Weinberg proportions



# Hardy-Weinberg proportions

6000 SNPs from HapMap (YRI)



YRI:  
Yoruba in  
Ibadan,  
Nigeria  
 $N = 60$

# Test of Hardy-Weinberg proportions

Genotype	AA	AG	GG	Sum
Observed ( $O_i$ )	968	1914	1434	4316
Expected ( $E_i$ )	$p^2N$	$2pqN$	$q^2N$	4316
Expected ( $E_i$ )	858.58	2132.84	1324.58	4316

## Test

$$\chi^2 = \sum (O_i - E_i)^2 / E_i$$

$$= (968 - 858.58)^2 / 858.58 \\ + (1914 - 2132.84)^2 / 2132.84 + (1434 - 1324.58)^2 / 1324.58$$

$$= (109.42)^2 / 858.58 \\ + (-218.84)^2 / 2132.84 + (109.42)^2 / 1324.58 \\ = 45.44$$

# Test of Hardy-Weinberg proportions

## Degrees of freedom

$H_0$ Genotypes	$A_1A_1$	$A_1A_2$	$A_2A_2$	Sum	Degrees of freedom
	$P_{11}$	$P_{12}$	$P_{22}$	1	2

$H_1$ Alleles	$A_1$	$A_2$		
	$p$	$q$	1	1
Genotypes	$p^2$	$2pq$	$q^2$	

---

Test 1  
(difference in degrees of freedom between hypotheses)

If  $X^2 > 3.84$ , the test is significant at the 5% level (df = 1).

# Test of Hardy-Weinberg proportions

Genotype	AA	AG	GG	Sum
Observed ( $O_i$ )	968	1914	1434	4316
Expected ( $E_i$ )	$p^2N$	$2pqN$	$q^2N$	4316
Expected ( $E_i$ )	858.58	2132.84	1324.58	4316

## Test

$$\chi^2 = \sum (O_i - E_i)^2 / E_i$$
$$= 45.44$$

R

```
> pchisq(45.44, df=1, lower.tail=FALSE)
[1] 1.57387e-11
```

# Interpreting a test for Hardy-Weinberg proportions

Accepting a hypothesis that an observed genotype distribution is in accordance with Hardy-Weinberg proportions,

**does NOT indicate that the population  
is in a Hardy-Weinberg equilibrium**



Hardy-Weinberg proportions,  
but **NO** Hardy-Weinberg equilibrium

$A_1A_1$

$A_1A_2$

$A_2A_2$

Sterile

Fertile

Sterile

1/4

1/2

1/4

# Deviations from Hardy-Weinberg proportions

Characterising deviations from Hardy-Weinberg proportions:

Definition of inbreeding coefficient

$$\begin{aligned} F &= (H_e - H_o)/H_e & H_o &\text{ Observed freq. of heterozygotes} \\ &= (2pq - P_{12})/(2pq) & H_e &\text{ Expected freq. of heterozygotes} \\ &= (2f_A f_a - f_{Aa})/(2f_A f_a) & &\text{ (Nielsen \& Slatkin notation)} \end{aligned}$$

therefore,

$$P_{12} = 2pq(1 - F)$$

With inbreeding or population admixture ( $F > 0$ ):

deficiency of heterozygotes compared to HW  
and excess of homozygotes compared to HW

# Deviations from Hardy-Weinberg proportions

Characterising deviations from Hardy-Weinberg proportions:

	Homozygotes	Heterozygotes
--	-------------	---------------

Inbreeding	↑	↓
------------	---	---



Population admixture	↑	↓
----------------------	---	---



Selection	↑↓	↑↓
-----------	----	----



Sex-linked polymorphism	↑	↓
-------------------------	---	---



Assortative mating	↑↓	↑↓
--------------------	----	----



# Deviations from Hardy-Weinberg proportions

Genotype	AA	AG	GG	Sum
Observed ( $O_i$ )	968	1914	1434	4316
Expected ( $E_i$ )	$p^2N$	$2pqN$	$q^2N$	4316
Expected ( $E_i$ )	858.58	2132.84	1324.58	4316

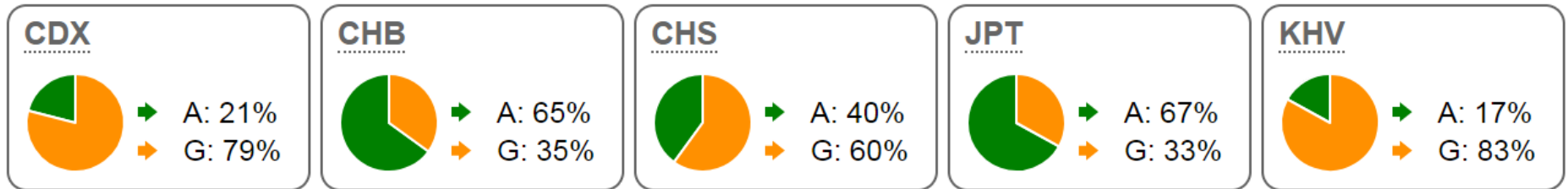
## SNP in East Asia population (EAS)

$$F = (2132.84 - 1914) / 2132.84 \\ = 0.11$$

What is the reason for the excess of homozygotes?

# Deviations from Hardy-Weinberg proportions

## EAS subpopulations



Three Chinese populations

Japan

Vietnam

Wahlund effect

# Hardy-Weinberg principle with sex linkage

Females	$X^A X^A$	$X^A X^a$	$X^a X^a$
Males	$X^A Y$		$X^a Y$
Allele frequencies	$X^A: p$ $X^a: q$		

Assumption: same allele frequencies in both sexes

# Hardy-Weinberg principle with sex linkage

Assumptions: Random union of gametes

Mendelian segregation

			Eggs	
			$X^A$	$X^a$
			$p$	$q$
Sperm	$\frac{1}{2}$	$X^A$	$p^2$	$pq$
		$X^a$	$qp$	$q^2$
	$\frac{1}{2}$	$Y$	$p$	$q$

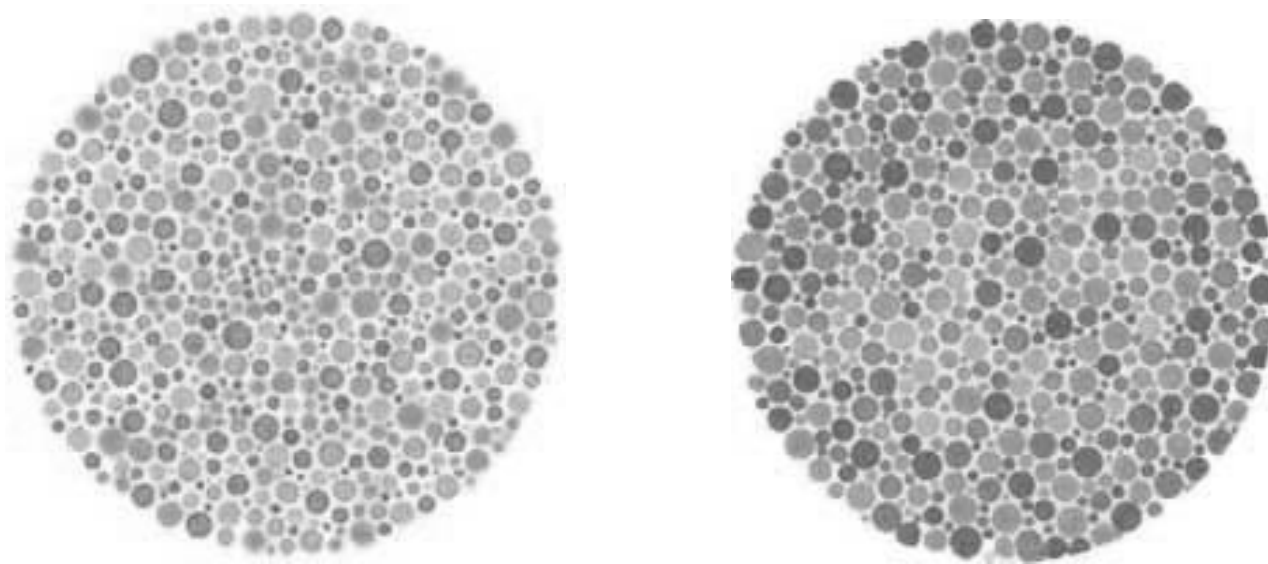
# Hardy-Weinberg principle with sex linkage

Females	$\frac{1}{2}$	$X^A X^A$	$X^A X^a$	$X^a X^a$
		$p^2$	$2pq$	$q^2$
Males	$\frac{1}{2}$	$X^A Y$		$X^a Y$
		$p$		$q$



# Hardy-Weinberg principle with sex linkage

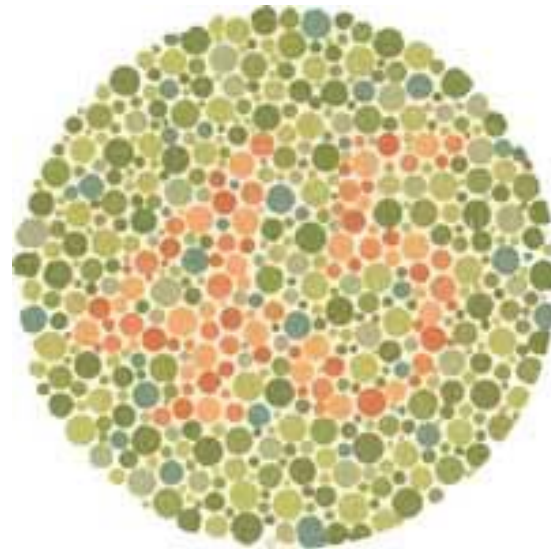
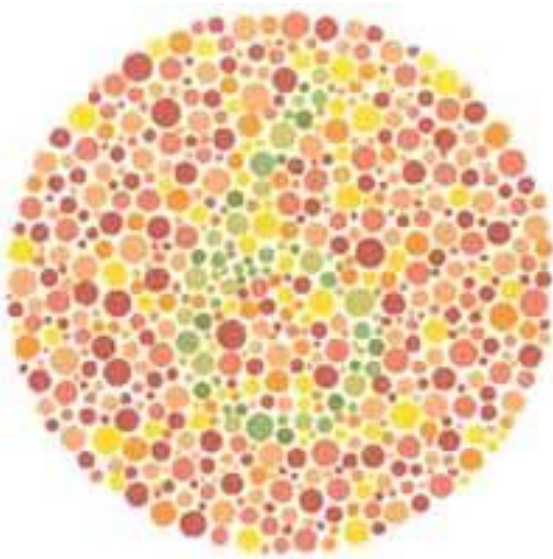
Which numbers can you see?



Example: Colour blindness is recessive in females

# Hardy-Weinberg principle with sex linkage

Which numbers can you see now?



Example: Colour blindness is recessive in females

# Hardy-Weinberg principle with sex linkage

Females	$X^A X^A$	$X^A X^a$	$X^a X^a$
	$p^2$	$2pq$	$q^2$
	0.846	0.147	0.0064
Males	$X^A Y$		$X^a Y$
	$p$		$q$
	0.92		0.08

Assumption: same allele frequencies in both sexes

Example: Colour blindness is recessive in females

$q$  in Denmark is  $\approx 0.08$

# Hardy-Weinberg principle with dominance interaction among alleles

Genotype	TT	Tt	tt
Phenotype	T-		tt
Number	$N_{T-}$		$N_{tt}$
	$N$ (total)		

$$\begin{array}{ccc}
 p^2 & 2pq & q^2 \\
 \hline
 p^2 + 2pq & & q^2
 \end{array}$$

Assuming HW:  $q = \sqrt{N_{tt}/N}$

# Analysis of genetic variation at a locus in a natural population

In a plant species the color of the flowers is determined by two alleles, one is dominant and the other recessive.

In a population of this species, 75% of the individuals carry red flowers and 25% white flowers.

What is the frequency of the allele coding for white flowers?

# Exercise 1 to 3

# Graphical test for multiple diallelic loci

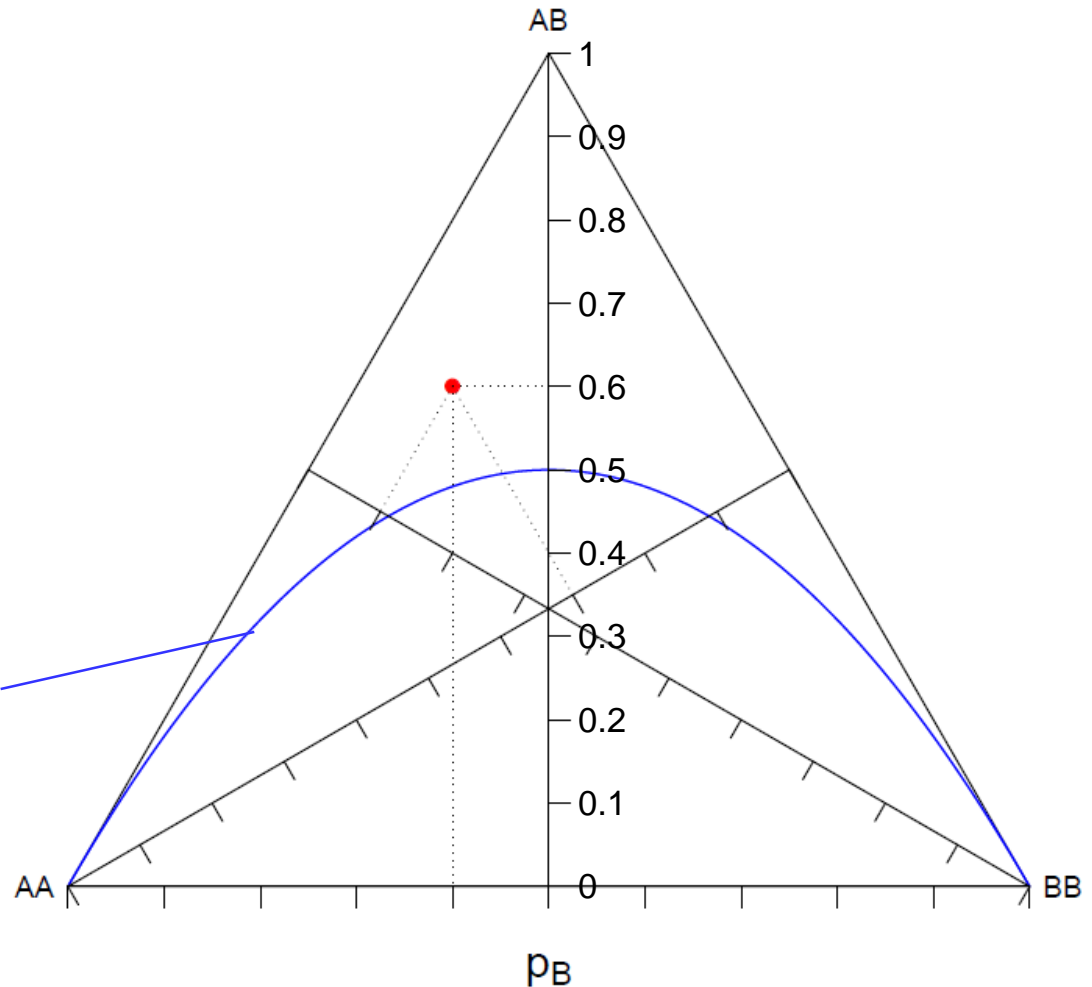
De Finetti diagram

$$f_{AA} = 0.3$$

$$f_{AB} = 0.6$$

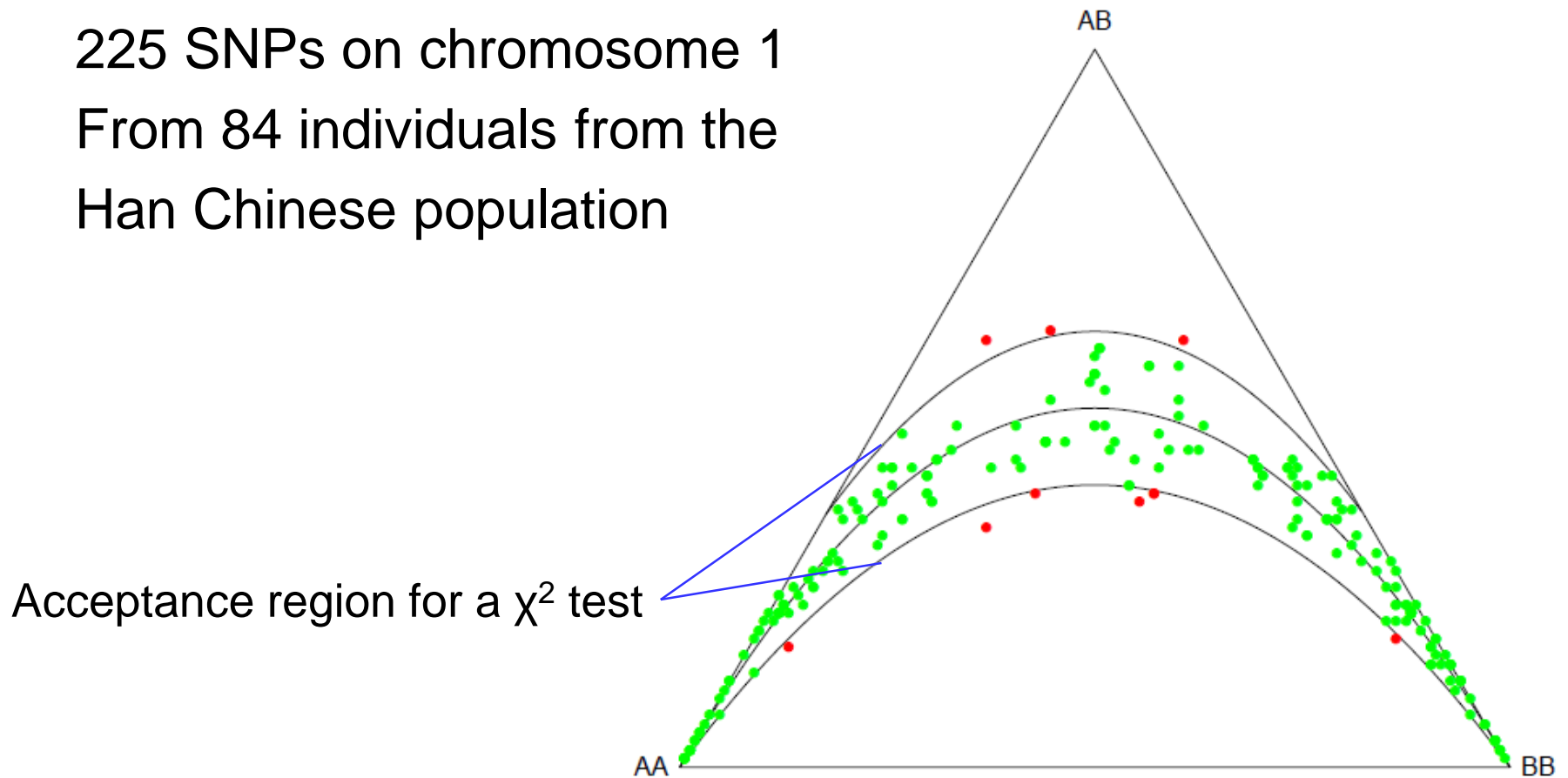
$$f_{BB} = 0.1$$

Hardy-Weinberg proportions



# Graphical test for multiple diallelic loci

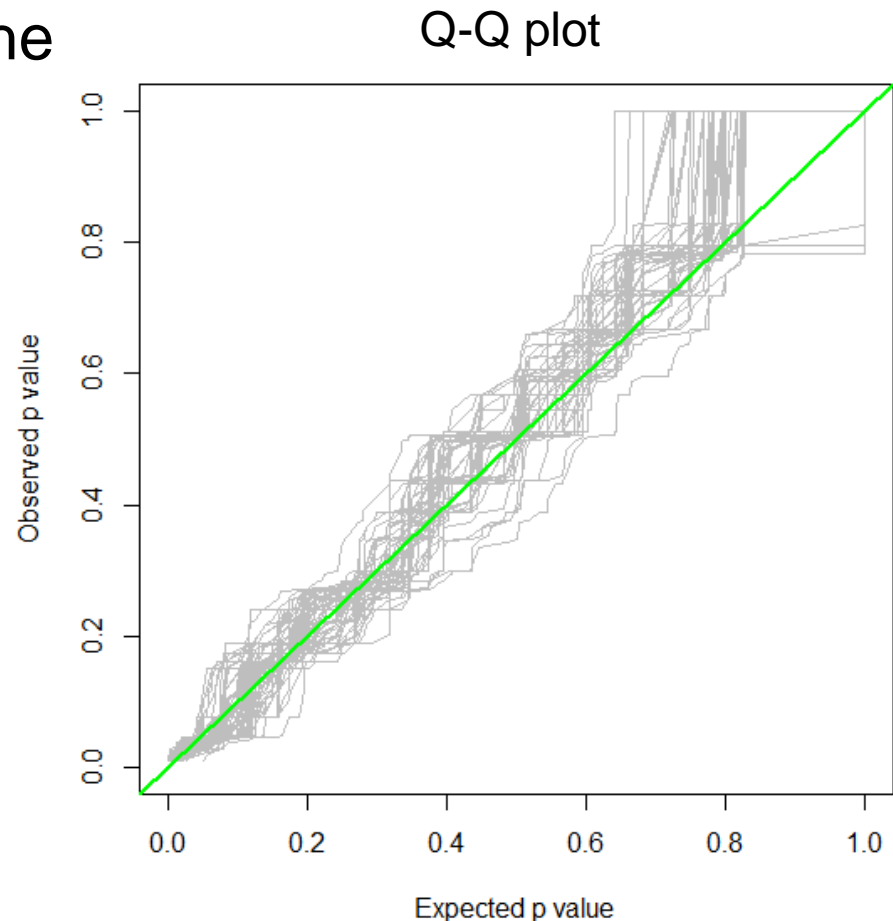
225 SNPs on chromosome 1  
From 84 individuals from the  
Han Chinese population





# Graphical test for multiple diallelic loci

225 SNPs on chromosome 1  
From 84 individuals from the  
Han Chinese population



# The Hardy-Weinberg principle

Concluding remarks

No models are correct.

But some are **extremely** useful.

We relax the assumptions

- Infinite population

- No mutation

- No migration

- No selection