Principles of Population Genetics

Leo P ten Kate, MD, PhD
Em. Prof. of Clinical Genetics
VU University Medical Center
Amsterdam, the Netherlands

Training Course in Sexual and Reproductive Health Research
Geneva 2012

Population Genetics

Population Genetics concerns the frequency distributions of genotypes and alleles in populations, as well as the determinants and possible fate of these distributions

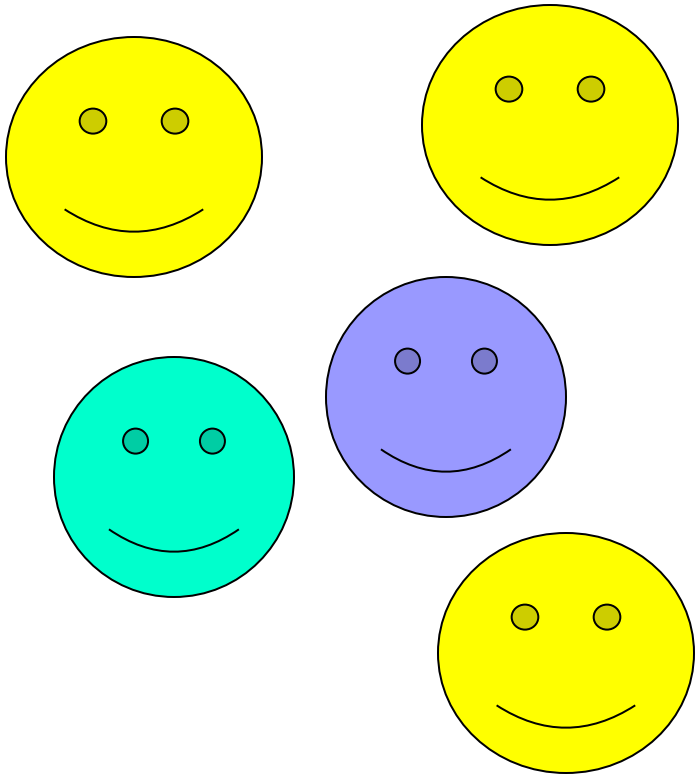
Content

1. Frequencies of genotypes and alleles
2. Populations
3. Frequencies in the basic idealized population
4. Some applications of Hardy-Weinberg equilibrium
5. Small populations
6. Non-random mating
7. Selection
8. Relaxed selection
9. Summary

1

Frequencies of genotypes and alleles

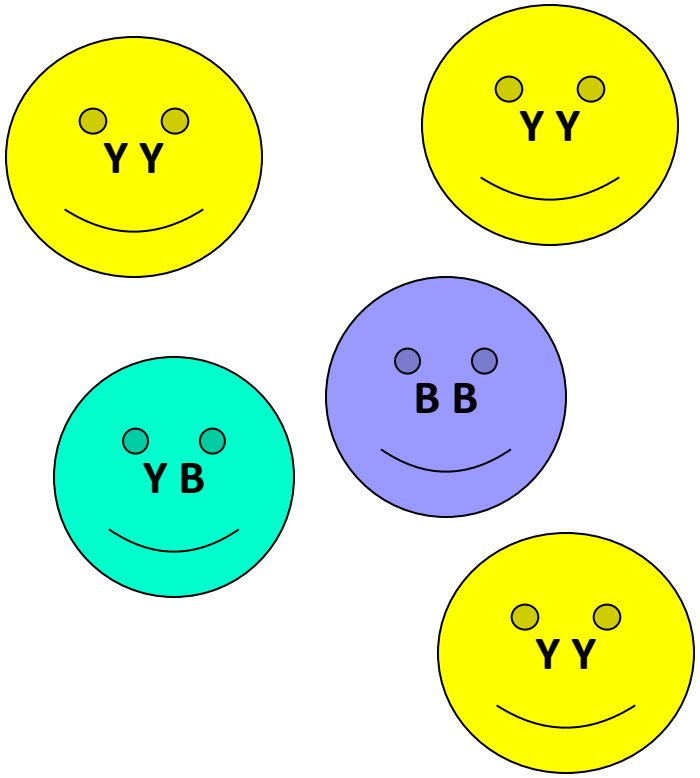
Frequencies of genotypes and alleles



In this example we have 5 persons, and 3 phenotypes: yellow, green and blue

We assume that the 3 phenotypes correspond to the 3 possible genotypes for a gene with 2 co-dominant alleles, Y and B (yellow and blue)

Frequencies of genotypes and alleles



- Frequencies of genotypes:
  - YY: 3/5 60%
  - YB: 1/5 20%
  - BB: 1/5 20%
- Frequencies of alleles:
  - Y: 7/10 70%
  - B: 3/10 30%

Frequencies of genotypes and alleles

- Frequency of genotypes:
  - Numerator: count people with particular genotype
  - Denominator: count all people
- Frequency of alleles:
  - Numerator: count all particular alleles
    - Some people have 1 of them, others 2 or none
  - Denominator: count all people twice

Limitations of the previous example

- Gene was bi-allelic, i.e. autosomal locus
  - In X-linked genes
    - Numerator: add count of allele in males and females
    - Denominator: all males + twice all females
  - In Y-linked genes
    - Numerator: count of particular allele in males
    - Denominator: all males once.

- Alleles were co-dominant
  - Principle applies also to dominant and recessive alleles, but counting will be more difficult (see later)

Guidance for students

Reasoning in population genetics frequently uses algebraic notations and expressions. This applies not only to text books but, also to the present presentation. Without algebraic notations texts would have to expand 10 or 100 times, and reading would become cumbersome.

For those students who lack trust or competence in reading algebraic expressions, there are two options:
(1) Find someone who can explain the algebra used
(2) Skip the algebra and accept its conclusions

Of course a choice for option 1 is recommended.

Allele frequency can be defined as

- The proportion of a particular allele among all alleles at one locus in a population
- The probability to find a particular allele if alleles at that locus are sampled at random from a population
- Note: allele frequencies are often called gene frequencies. Why is this a misnomer?

Indicating allele frequencies

- Allele frequencies are generally indicated by small letters
- If there are only two alleles, the frequency of the more frequent allele is indicated by *p,* while the frequency of the less frequent allele is indicated by *q*
- Note: $p + q = 1$
- If there are > 2 alleles, other letters can be used, e.g. *p*, *q*, *r*, *s*, …
- Note: $p + q + r + s + \ldots. = 1$

From genotype frequencies to allele frequencies

- While allele frequencies usually are indicated by small letters, we will use capitals for genotype frequencies, e.g. D, E and F for genotypes YY, YB and BB, respectively.

  Note that D + E + F = 1

- Accordingly, the allele frequency *p* for allele Y is (2xD + E)/2 and the allele frequency *q* for allele B is (E + 2xF)/2.

  Note that [(2xD +E)/2] + [(E + 2xF)/2] = 1

Allele frequency predicts gamete frequency

Genotype	AA	Aa	aa	total
Number*	300	200	100	600

* Arbitrary chosen

Allele frequency predicts gamete frequency

Genotype	**AA**		**Aa**		**aa**		**total**	
Number*	300		200		100		600	
Alleles	A	A	A	a	a	a	A	a

* Arbitrary chosen

Allele frequency predicts gamete frequency

Genotype	**AA**		**Aa**		**aa**		**total**	
Number*	300		200		100		600	
Alleles	A	A	A	a	a	a	A	a
Number **Frequency**	300	300	200	200	100	100	800 **2/3**	400 **1/3**

* Arbitrary chosen

Allele frequency predicts gamete frequency

Genotype	**AA**		**Aa**		**aa**		**total**	
Number*	300		200		100		600	
Alleles	A	A	A	a	a	a	A	a
Number	300	300	200	200	100	100	800	400
Frequency							**2/3**	**1/3**
Gametes (relative)	800			400			2/3	1/3

* Arbitrary chosen

2

Populations

There are 2 types of populations

- Real populations, such as the inhabitants of your country.
- Idealized populations.
  - Used as models to understand how genotype and allele frequencies affect each other; are affected by different parameters; and develop over time.
  - Used in comparisons with real populations to understand real population better.
  - One basic idealized population, and many other idealized populations which deviate in one or more characteristics from the basic one.

Characteristics of the basic idealised population

- The population is large
  - The influence of change can be ignored
- Mating is at random
- There is no migration in or out
- There are no new mutations
- There is no differential selection
  - Each genotype has equal chance of survival and reproduction

Real populations

- Real populations may deviate more or less from the basic idealized population.
- The deviation may affect only one gene. Allele frequencies of that gene for instance may increase by new mutations, or decrease by selection against carriers or patients. For some diseases both apply.
- Random mating will seldom be reached completely. Many populations are sub-structured by mating patterns.

3

Allele and genotype frequencies in the basic idealized population

What happens to our previous hypothetical population under basic idealized conditions?

- Genotype frequencies were 300, 200 and 100 for AA, Aa and aa, respectively
- Allele, and also gamete frequencies were 2/3 and 1/3 for A and a, respectively
- Random mating of individuals (male/female) equals random fusion of gametes (sperm/egg cells)
- So the frequency of genotypes in the next generation will be
  - 2/3 x 2/3 = 4/9 for AA; which is 267 if the size of the population is stable
  - 1/3 x 1/3 = 1/9 for aa; which is 67 if the size of the population is stable
  - 2/3 x 1/3 x2 = 4/9 for Aa; which is 267 if the size of the population is stable
- Note: 4/9 + 1/9 + 4/9 = 1

Generalisation

Gametes	Female			
Male	Alleles		**A**	**a**
		Fre-quencies	p	q
	A	P		
	a	q		

Generalisation

Gametes	Female			
Male	Alleles		**A**	**a**
		Fre-quencies	p	q
	A	P	p^2	pq
	a	q	pq	q^2

Generalisation

- So, after one generation, the ratio of genotype frequencies changed
  - from 1/2 : 1/3 : 1/6
  - to 4/9 : 4/9 : 1/9

- The ratio is now $p^2 : 2pq : q^2$, with $p^2 + 2pq + q^2 = 1$

What happens in the next generation?

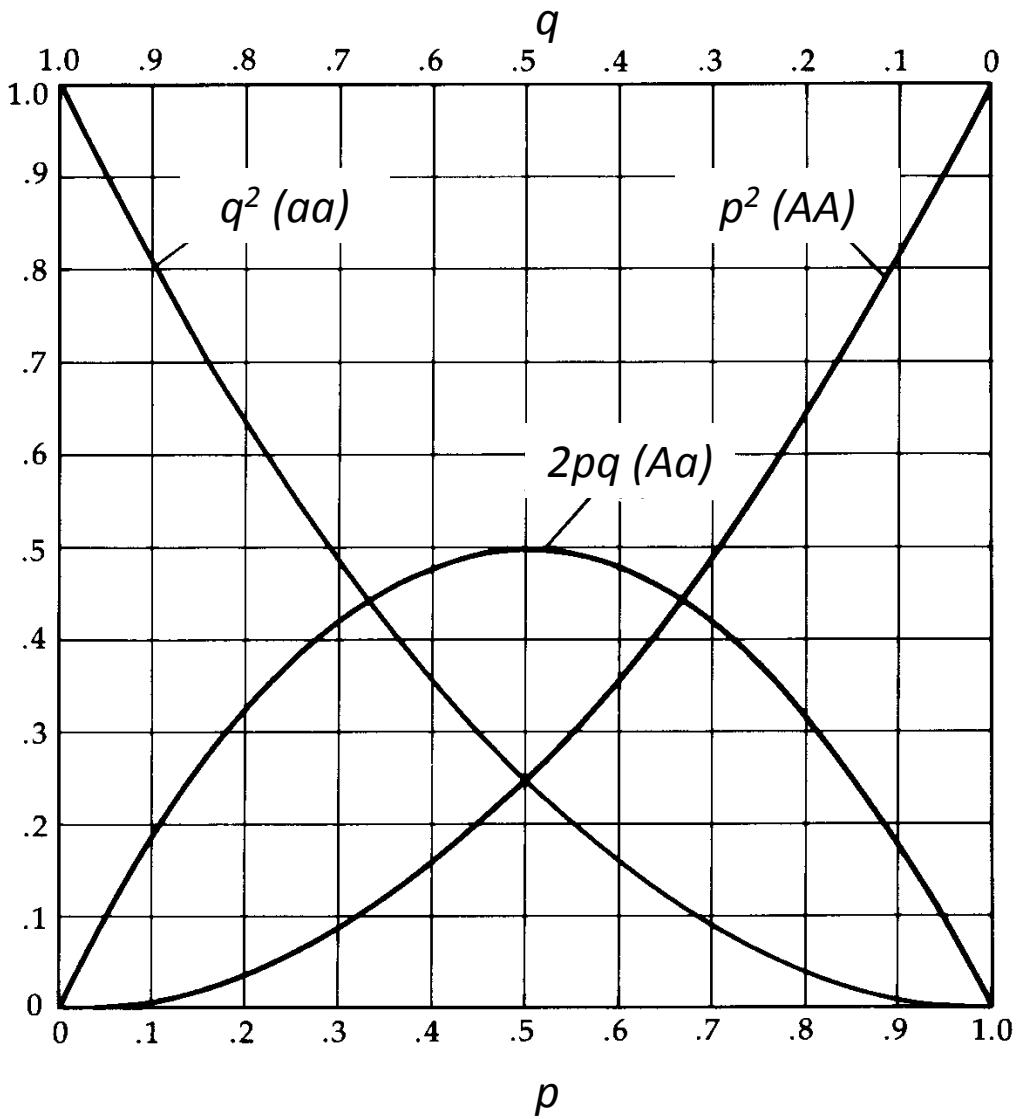
- The genotype frequencies in the parental generation are now: p^2 (AA), 2pq (Aa), and q^2 (aa), respectively.
- Gamete frequencies are
  - For A: $p^2 + pq = p(p+q) = p(1) = p$
  - For a: $q^2 + pq = q(q+p) = q(1) = q$
- So, genotype frequencies in the next generation again will be p^2 (AA), 2pq (Aa), and q^2 (aa), respectively.
- Once genotype frequencies can be explained by allele frequencies in the parental generation, they will not change anymore.
- Genotype and allele frequencies are in equilibrium (so called **Hardy-Weinberg equilibrium**).

Hardy and Weinberg formulated this equilibrium relationship, independent from each other, in 1908

Geodfrey Harold Hardy,
England, 1877- 1947
Mathematician

Wilhelm Weinberg
Germany, 1862-1937
Physician

Anyone variable (either allele or genotype frequency) is sufficient to predict the others



Other types of equilibrium

Other types of equilibrium are:

- A **mutation-selection** equilibrium, in which the loss of alleles by selection is compensated for by new mutations. Examples are many monogenic diseases.
- A **selection-selection** equilibrium, in which selection against one genotype is compensated for by selection against another genotype. An example is sickle cell anaemia.
- Note that in these cases the conditions for the basic idealised population (especially no mutation and no differential selection) are not met.

4

Some applications of HW-equilibrium

The frequency of carriers for an autosomal recessive disease

- If one knows the population frequency of an autosomal recessive disease, one can calculate the population frequency of carriers of this disease.

- The disease frequency is q^2 . So the allele frequency is the square root of it, i.e. q, the frequency of the normal allele (1-q), and the carrier frequency is $2pq$, or $2(1-q)q$.

- This is important knowledge for counselling family members of patients with autosomal recessive diseases.

Is carrier frequency $2pq$ or $2q$?

- For rare diseases p is almost 1. So $2pq \approx 2q$.
- For lethal diseases, in which patients do not contribute to the next generation, all patients have two carrier parents. If the carrier frequency in the population is C, the frequency of carrier couples will be C^2, and the frequency of the disease will be $C^2/4$. As $C^2/4 = q^2$, $C^2 = 4q^2$, and therefore $C = 2q$.
- For frequent, non-lethal autosomal recessive disorders the carrier frequency remains $2pq$.
- Note: lethality is a violation of the conditions of the basic idealized population.

Checking for HW equilibrium

- Suppose in a random sample of a big city we observe the following distribution of genotypes

Genotype	Observed
1,1	82
1,2	36
2,2	82
total	200

- Is this population in HW equilibrium?

Checking for HW equilibrium

- From the observed genotype distribution we find

  p = (2x82 + 36)/2x200 = 0.5. Likewise q = 0.5. So the expected distribution is as shown:

Genotype	Observed	Expected
1,1	82	50
1,2	36	100
2,2	82	50
total	200	200

- The observed values deviate significantly from the expected ones under HW equilibrium. How come?

Checking for HW equilibrium

- One possibility is that the population is a mixture of two or more populations, each one of which is in HW-equilibrium itself.

Genotype	Observed	Pop. 1	Pop. 2
1,1	82	1	81
1,2	36	18	18
2,2	82	81	1
total	200	100	100

- Here we have two populations of equal size, but with unequal allele frequencies (p = 0.1 in Population 1, and p = 0.9 in Population 2.

Checking for HW equilibrium

Some other possibilities:

Genotype	Observed
1,1	82
1,2	36
2,2	82
total	200

- heterozygotes move out of town
- heterozygotes die younger
- positive assortative mating
  (like draws to like)
- selection was not random after all

5

Small populations

The smaller the population is, the larger are the sampling effects

- **Genetic drift** refers to change fluctuations in allele frequency as a result of random sampling among a relatively small number of gametes.
- This process eventually leads to **loss** of alleles (allele frequency is zero) or to **fixation** of alleles (allele frequency is 1).

Simulation of 24 populations of size N = 9, during 19 generations, starting with allele frequency 0.5. Result: Fixation in 7, and complete loss in 6 populations. Hartl, 1980

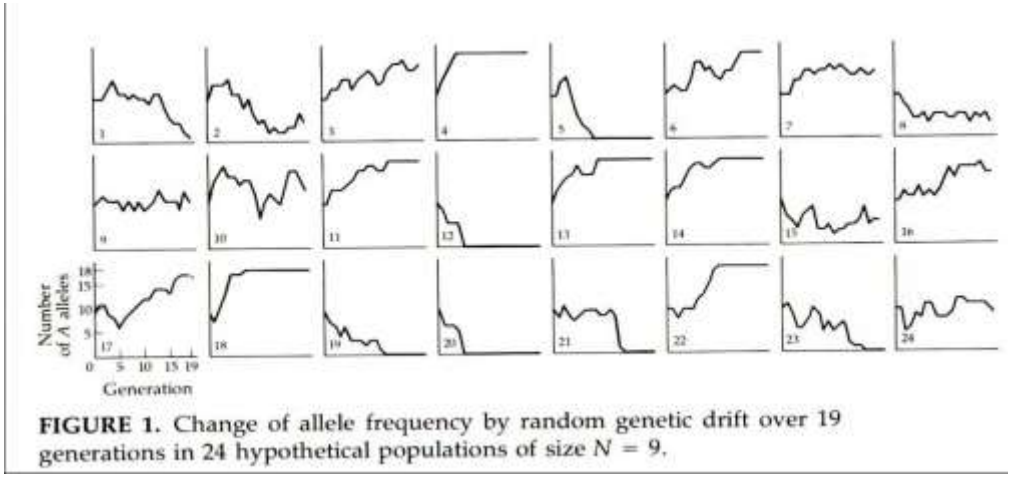


FIGURE 1. Change of allele frequency by random genetic drift over 19 generations in 24 hypothetical populations of size $N = 9$.

Founder effect

- If a small sample of a larger population becomes isolated from that population, some of its allele frequencies can deviate significantly.
- Genetic drift may strengthen or diminish founder effect.
- The high frequency of a genetic disease in a population frequently is due to a founder effect.
- Events producing founder effect are migration and bottlenecks.

6

Non-random mating

Non-random mating

- Positive assortative mating
  - Mates are phenotypically more similar than expected by chance
- Negative assortative mating
  - Mates are phenotypically more dissimilar than expected by chance
- Consanguinity
  - Mates are relatives
- Population substructure

Assortative mating

- Examples from complex traits are height and school performance.
- An example from monogenic conditions is deafness.
- Assortative mating changes genotype frequencies.
- Assortative mating does not change allele frequencies.

Consanguinity

- Increases the probability of inheriting alleles identical by descent.
- This probability is represented by the inbreeding coefficient (F).
- F is larger as the parents are more closely related.
- Some typical values of F are
  - 1/8 for uncle-niece
  - 1/16 for first cousins
  - 1/64 for second cousins
- The probability that two alleles are not identical by descent is (1-F).

Consanguinity

- The probability for becoming homozygous is

  $$Fq + (1-F)q^2$$

- If F = 0 (no inbreeding) $Fq + (1-F)q^2$ equals q^2
- If the parents of a child with an autosomal recessive disease are related, the chance that the disease in this child is not due to the consanguinity of the parents, is

  $$(1-F)\, q^2/[Fq + (1-F)q^2]$$

- If a recessive disease is caused by homozygosity or compound heterozygosity of several alleles, q in the above equations stands for the total frequency of disease associated alleles.

7

Selection

Items

- Uncompensated selection.
- Mutation frequency in case selection is compensated by new mutations.
- Selection coefficients in case of balancing selection against homozygotes.

Uncompensated selection

Uncompensated selection will lead to loss of alleles -> decreasing frequency

Example: autosomal recessive disease; complete selection against patients

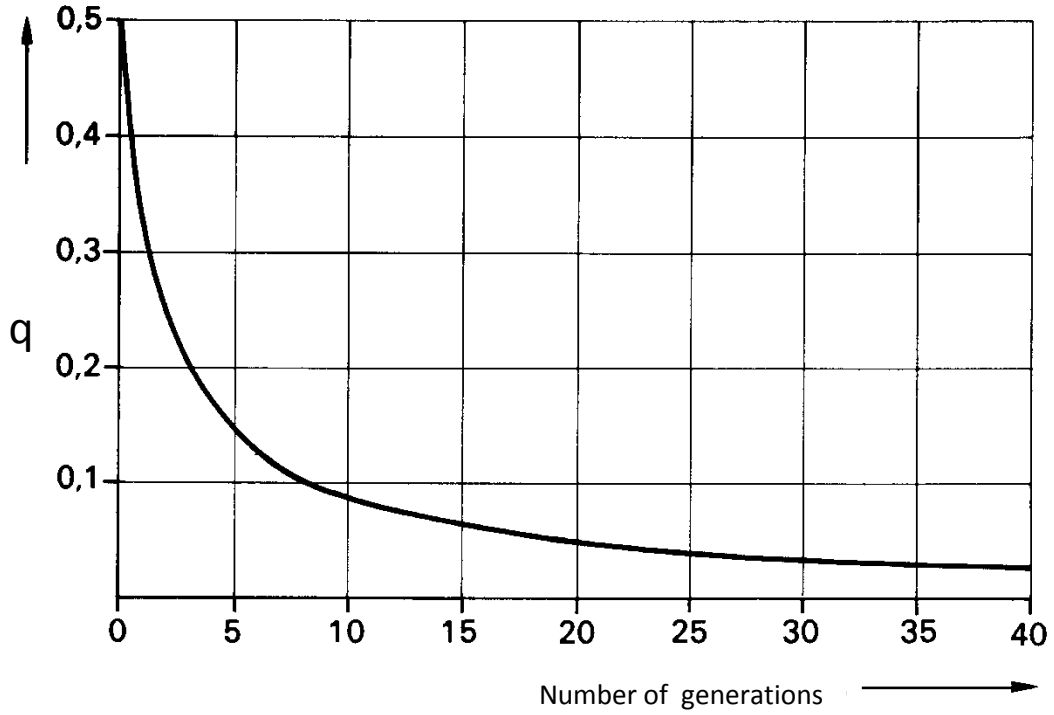
Generation 0	Genotypes			total
	AA	Aa	aa	
Frequency before selection	p^2	2pq	q^2	1
Fitness	1	1	0	
Frequency after selection	$p^2/(p^2+2pq) = p/(1+q)$	$2pq/(p^2+2pq) = 2q/(1+q)$		1

Allele frequency in the next generation: $q_1 = q_0 / (1+q_0)$

Allele frequency after n generations will be $q_n = q_0 / (1+nq_0)$

Complete uncompensated selection against autosomal recessive diseases

In the end, uncompensated selection will end in the disappearance of alleles and diseases caused by such alleles.



Mutation/selection equilibrium

- Allele and genotype frequencies remain constant if the loss of alleles by selection, is compensated by new mutations.
- For lethal autosomal recessive diseases equilibrium is reached when the mutation frequency, μ, equals the disease frequency, q^2.
- In general $\mu = s.q^2$, with s being the selection coefficient.
- For the more frequent autosomal recessive diseases, such as cystic fibrosis or sickle cell anemia, a mutation/selection equilibrium is not plausible, as it would require improbable high mutation frequencies.

Equilibrium by selection against both types of homozygotes

- Severe selection against the less frequent homozygote can be balanced by slight or moderate selection against the more frequent homozygote.
- This situation is also known as heterozygote advantage.
- The classical example is sickle cell anaemia
  - Homozygotes with the Q6V mutation in the β-globin gene are severely affected.
  - Homozygotes with the normal allele are more susceptible to malaria than the heterozygotes.

How much selection against AA is needed to counterbalance complete selection against aa?

Genotypes	**AA**	**Aa**	**aa**
Before selection	p_0^2	$2p_0q_0$	q_0^2
Fitness	1-s	1	0
After selection	$\frac{p_0^2(1-s)}{p_0^2(1-s)+2p_0q_0}$	$\frac{2p_0q_0}{p_0^2(1-s)+2p_0q_0}$	

So, $q_1 = \frac{p_0q_0}{p_0^2(1-s)+2p_0q_0} = \frac{q_0}{p_0(1-s)+2q_0}$

At equilibrium $q_1 = q_0 = q$, and $p_1 = p_0 = p$. So, $p(1-s) + 2q = 1$

So, $s = q/(1-q)$

How much selection against AA is needed to counterbalance complete selection against aa?

Lethal autosomal recessive disease	Prevalence at birth	Frequency of allele a	Selection coefficient (against AA)
A	1 : 2,500	1 : 50	0.0204
B	1: 10,000	1 : 100	0.0101
C	1 : 40,000	1 : 200	0.0050

8

Relaxed selection

How long will it take to double the disease frequency of a genetically lethal autosomal recessive disease from onset of effective screening and therapy on

- Assumption: Before onset of effective screening and therapy there was a mutation/selection balance. So $\mu = q^2$
- Example: PKU
- Prevalence at birth before onset of effective screening and therapy was 1 : 15,000 (So, $q_0 = 1/122$)
- From onset of effective screening and therapy on mutations are no longer balanced by selection. So allele frequency will rise
- Doubling of disease frequency means 1:7,500; So $q_n = 1/87$

How long will it take to double the disease frequency of a genetically lethal autosomal recessive disease from onset of effective screening and therapy on

- $p_1 = p_0 - \mu.p_0 = (1 - \mu).p_0$
- $p_2 = p_1 - \mu.p_1 = (1 - \mu).p_1 = (1 - \mu).(1 - \mu).p_0 = (1 - \mu)^2.p_0$
- $p_n = (1 - \mu)^n.p_0$
- $(1-q_n) = (1 - \mu)^n.(1-q_0)$
- $(1 - \mu)^n = (1-q_n)/(1-q_0)$
- $n.\log(1 - \mu) = \log[(1-q_n)/(1-q_0)]$
- n = 51 generations (1000 - 1500 year)

9

Summary

Summary

- Allele frequencies can be determined from genotype frequencies.
- Frequencies of alleles in gametes correspond to allele frequencies in the population.
- Random mating equals random fusion of gametes.
- On conditions genotype frequencies AA, Aa and aa are in the proportion of p^2 to $2pq$ to q^2 (Hardy-Weinberg proportions).
- The conditions are: a very large population, random mating, and absence of migration, mutation and selection.

Summary

- The effect of violation of one or more conditions can be calculated by means of relatively easy algebra.
- Apart from HW other mechanisms may result in equilibrium, e.g. heterozygote advantage or counteraction of mutation and selection.
- Some relatively frequent applications are
  - Calculating carrier frequency from autosomal recessive disease frequencies.
  - Calculating risks in case of consanguinity.
  - Checking whether observations answer HW expectations.