Population genetics by [Knud Christensen](#)

Division of Animal Genetics

Introduction

The present genetics notes are produced as a substitute for '*Veterinary Genetics*' by FW.Nicolas, Oxford University Press, 1989. This book was not available after 1995.

The notes are produced for a course for veterinary students. Therefore, efforts have been made to adjust the notes for students with a biological background, and at the same time supply a minimal set of formulas to describe the relationship between practical observations and genetic theory. In addition to the description of traits with simple Mendelian inheritance, the description of the genetics for traits (diseases) with multifactoriel aetiology has also been emphasized, here the application of correct breeding plans make it possible to significantly lower the frequency of the disease.

The genetics notes are available on the www server, which can be reached from the address: www.ihh.kvl.dk/htm/kc/popgen/genetics/genetik.htm. Both an English and a Danish version is available. The online voice version includes a slide show with more than 350 slides. The voice in the voice version

belongs to Anne Asp Poulsen who study English at the Copenhagen University, she also gave comments on the English style in the text and figures.

The online notes include a number of links (underlined in the text) to other servers and to extended calculation examples and data programs (applets).

There is no independent exercise section yet in the English version. But to each applet there is an example and one or more exercises for solution.

Ass. professor Peter Sestoft has advised me while producing the applets.

Some students which had the important background of being in the process of learning the topic have commented on the text. Any proposals for additional improvements are welcome.

2. English edition, May 2003; with revision 2014 Knud Christensen

References.

Christensen, L.G., *Husdyravl - teori og praksis,* DSR Forlag, 1999.

Falcorner,D.S. & Mackay T., *Introduction to Quantitative Genetics,* Longman Scientific & Technical, 1996.

Nicolas, F.W., *Veterinary Genetics*, Oxford University Press, 1989.

Nicolas, F.W., *Introduction to Veterinary Genetics*, Oxford University Press, 1996.

Samuel, M., *Statistics for the life science*, Prentice-Halls,1989.

Strachan, T. & Read A.P., *Human Molecular Genetics*, pp.597, John Wiley & Sons, 1996.

Download Population Genetics, Pdf file

Chapter 1. Introduction, quantitative versus qualitative genetics

This chapter is meant as a brushup for terms such as genotype, phenotype and linear regression, as well as the introduction of new ones. This is only meant as an overview, so very detailed study should be avoided, as some of the terms are fairly abstract. After reading of chapters 2 to 8 read this chapter again.

1.1 Domestic animals in Denmark, quantitative traits

Danish populations of domestic animals/production per year, rounded numbers.

Figure 1.1.
Populations: Dogs 500, Cats 500, Horses 100, Sheep 100 and production per year Cattle 1000, Broilers 120000, Pigs 22000, Mink 10000
(all in 1000)
Fish 40, Butter 100, Cheese 300, Beef meat 300
(all in 1000 tons)
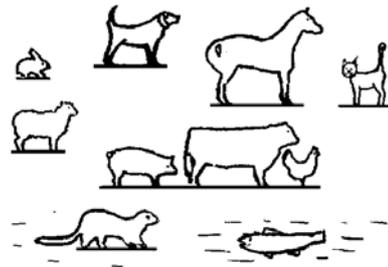
The summary list of the Danish animal production shows that quantitative traits have great significance for the magnitude and the economy of animal production. The joint production of domestic animals in Denmark amounts to around 50 billion Dkr per year. The main part is exported.
For world animal breeds, see Livestock, Oklahoma

1.2 Quantitative versus qualitative inheritance

There is a continuum of traits being inherited as a Mendelian trait with simple inheritance and traits having quantitative inheritance without well separated classes and with many genes involved.

Classification of traits in relation to mode of inheritance and environmental tolerance are shown in Figure 1.2. First there are the well known traits with simple Mendelian mode of inheritance. The trait with quantitative genetic inheritance is caused by segregation of many gene pairs, each with small effect. At the same time the trait is influenced by a lot of minor environmental effects.
Diseases will often be 'either/or traits' as the simple Mendelian traits. Cases in which the severity of the disease has a normal distribution can also be found. In many production diseases the disease only occurs when a genetically prone individual is exposed to adverse

Figure 1.2. Classification of traits in relation to mode of inheritance and environmental tolerance.

Trait	Mode of inheritance	Environmental effects	Diseases
Qualitative, Blood and serum types Enzyme defects	Simple Mendelian (monogene)	Little	
Quantitative, Stature height Intelligence Milk yield	Complex (polygene)	Moderate to great	
Threshold, Hare lib Pylero stenosis Many diseases	Complex (polygene)	Moderate to great	
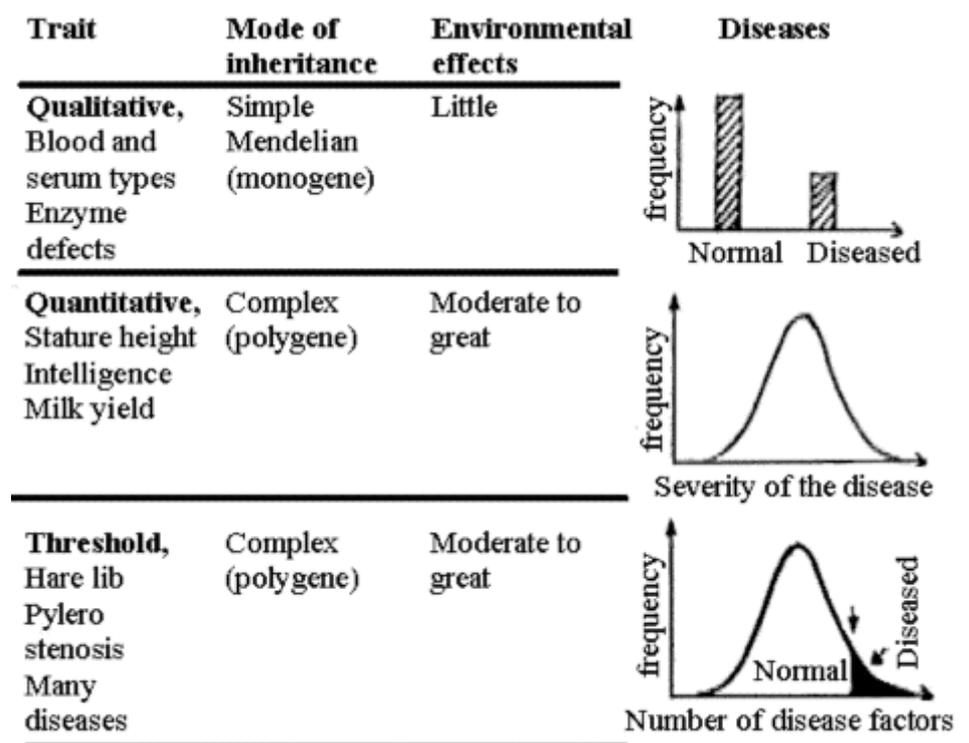
environmental effects. See Figure 1.2, produced by prof. emeritus Erik Andresen.

Figure 1.3 gives an illustration of how one or two Mendelian segregating gene pairs control the milk yield. For each A or B allele an individual has a yield increase of one kilogram. The alleles A and a have the same frequency in the distributions. For a realistic picture of the genetic background for milk yields, hundreds of gene pairs have to be involved. The milk yield has by selection been changed just as dramatically as the fat-% in the milk, as shown in section 1.4. To make this possible there has to be an effect from numerous gene pairs. In the present example using only two gene pairs, they could be fixed after one generation of selection.

Figure 1.3. Illustrating how one and two gene pairs can influence milk yields. In reality, numerous gene pairs have to be involved for selection to be carried out generation after generation without the genetic variation diapers when selecting .
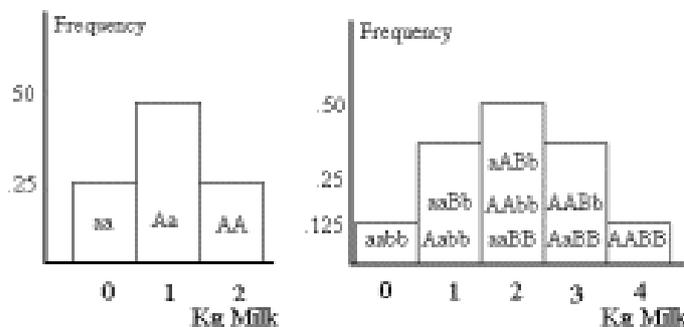
1.3 The terms phenotype, genotype and heritability

Most quantitative traits exhibit some degree of heritability. The heritability is evident when individuals, deviating positively or negatively from the average, also become offspring with deviation in the same trait in the same direction as their parents. There is a continuum of some traits, which is inherited with a simple Mendelian form and other traits with quantitative genetic inheritance without separate classes. The quantitative genetic inheritance is caused by the effect of many different genes, each with minor effect. The traits are also under the influence of environmental effects.

The similarity between related individuals is determined by the degree of heritability. The degree of heritability can be estimated statistically as a regression of offspring on average parents. The degree of heritability has values between 0 and 1. The degree of heritability at 0 corresponds to no similarity, and 1 corresponds to the highest possible similarity between parent and offspring. See right side of Figure 1.4.

The upper part of Figure 1.4 gives the relation formula for a trait or the entire individual the **Genotype** (all genes inherited from the parents) and **the phenotype** (appearance or what can be measured in the individual). The deviations of the phenotype from the genotype are caused by random environmental effects. The formulation was made by the geneticist Wilhelm Johannesen, employed by this university around year 1900. The formula was based on the size of bean seeds derived from beans with varying degrees of inbreeding.

Figure 1.4
Relation between phenotype, genotype and environment have been formulated by W. Johannesen based on the shown bean experiments.
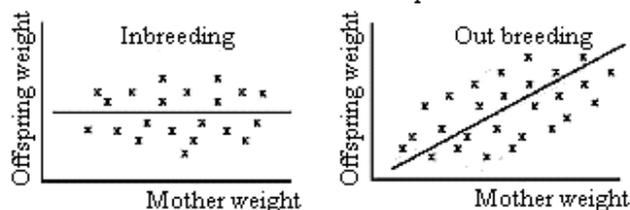
Figure 1.4 shows what Johannesen discovered: When the beans were 100 % inbred, which means that all beans were genetically similar, there was no relation between the weight of the parent bean and that of its offspring, i.e. the regression coefficient (b) of offspring on parents equaled 0. For outbred beans being

6

genetically different, there was a linear regression of b = 0.27. Which means that if a bean was 10 mg larger than average then the offspring was 2,7 mg larger than average.

For most traits in outbred populations there will be some similarities between related individuals. Figure 1.5 shows the twodimensional relationship between the average height of the vet. students and that of their parents. The figure is a printout from the SAS system by use of the procedure proc plot.

Figure 1.5 shows that there is a strong relationship between the height of the parent and that of their offspring. The relationship equals a degree of heritability of 60 % for the trait height in the human population, cf. the slope of the regression line which is b = 0.6.

The **Genotype** can only be observed when genetic variation occurs. This variation is equal to the part of the phenotype (the phenotypic variation) which can be passed down to the offspring (genotypic variation). The genotype constitutes respectively 27 % of the phenotypic variation for bean weight and 60% of human height. The

Figure 1.5
Relationship between the height of parent and offspring (vet. students). If the average height of the parents is 1 cm above the average, their offspring is 0.6 above the average. The relationship is caused by the fact that height, as a trait, is passed down with a heritability of 60 %. The symbol 1 is the boys and 2 is the girls.

environmental effects thus contribute to the rest of the variation. That is (100-27)= 73 % for bean weight and (100-60) = 40 % of the phenotypic variation for human height.

1.4 Effect of animal breeding (evolution)

The effect of animal breeding is shown in Figure 1.6. This gives an understanding of how the breeding work has affected the fat-% in the milk from the Danish Jersey dairy breed. Over the last 20 generations animals with the highest fat-% in the milk have been selected for breeding. The effect of this selection has been an increase by 0.1 units in

Figure 1.6.
Distribution curves for fat-% in the milk for HF (Holstein Frisian) and Jersey. For Jersey both year 1900 and 1990.

the fat-% per generation. Nothing indicates that it would impossible to continue for the next 20 generations with the same effect of selection to gain a higher fat content in the milk. Or whether it is desirable by selection to return to the starting point. The HF has not been selected for fat-% and has therefore been fairly stable with respect to the fat-% in the milk during the same period.

7

Figure 1.7 shows a drawing from *Science* Nov. 1990. The drawing indicates that evolution is a simple principle in which each animal adapts to the challenges of the environment. Evolution is not the individual animals' ability to adapt, but the fact that the best adapted animals give birth to a larger number of well adapted offspring, therefore new individuals which are even better adapted occurs in each new generation

Figure 1.7.

EVOLUTIONARY
THEORY MADE
SIMPLE
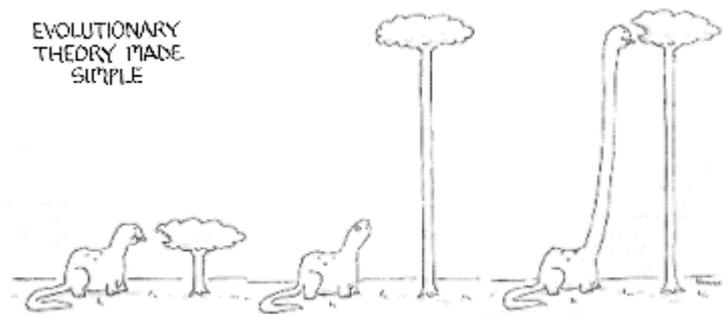
Both animal breeding and natural selection are slow processes. But with strong selection a population can change the average value with up to 10 % per generation, if the given trait has a high degree of heritability.

Results of continued selection for litter size in mice through 30 generations can be seen here. By selection in this period the litter size has been increased by 5 young per litter, from 9 to 14. This corresponds to an increase in litter size of 0.16 young per generation.

1.5 Qualitative traits, Mendelian genetics

Qualitative traits are characterized by segregation in the classical Mendelian ratios. An example of this is the coat colour gene in the Labrador Retriever which can occur either as black or as yellow. The yellow colour is recessive and the black is dominant. When a gene for yellow coat colour is mentioned it points at both a locus and an allele.

One can also talk about an Albumin locus, even the phenotypes cannot be seen directly. But as will be shown in the next chapter, polymorphism in the locus can be assigned by means of electrophoreses of serum samples separating the two albumin alleles. The word 'gene' should not be used for DNA polymorphism with more alleles in non-coding sequence. Instead it should be called a locus with more alleles. There is not always a sharp separation between gene and locus or between gene and allele in practice, therefore when gene frequency is mentioned in the next chapter, it will mean both proper genes as well as alleles in non-coding DNA sequences (loci).

Figure 1.8 Segregation of genetic variation (mikrosatellite) in a locus in a swine family detected by gel electrophorese of a PCR product. The boar Cup carries the alleles 209 and 195 and sow 400 carries the alleles 199 and 195 segregating in the offspring.

Figure 1.8 shows segregation of genetic polymorphism in a swine family. The polymorphism is in a mikrosatellite, 'S0002' locus, which has a (GT) repeat. It is detected by means of PCR followed by electroforeses of the product to separate the alleles. Most of the individuals have two bands, heterozygotes. Some individuals have only one band, homozygotes. The PCR analyses is carried out by Merete Fredholm.
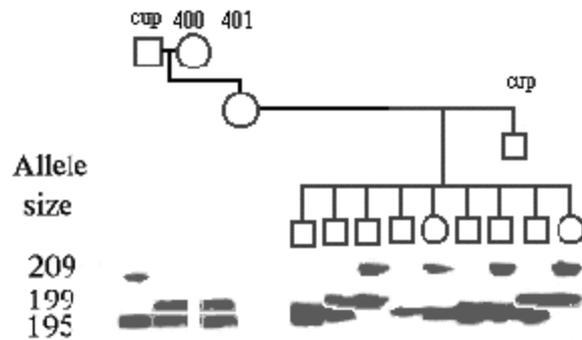
The boar Cup carries the alleles 209 and 195 and sow 400 carries the allele 199 and 195. Cup is both sire and grand sire of the litter. The dam of the litter number 401 has got the '195' allele from Cup and '199' allele from sow 400.
The shown litter is part of a bigger family, which will be part of a calculation example in the next chapter. For more information on the pig family click here.

1.6 Data base on Mendelian inheritance in domestic animals

To get an impression of Mendelian inheritance within domestic animals a visit to the Australian data base ANGIS is appropriate. It can be reached here. The data base mainly contains information on segregation of disease genes in populations of domestic animals. The data base is compiled by Nicolas who is also the author of the book *'Introduction to Veterinary Genetics'*. Below is shown a print of references from the data base on the albino gene in cattle. Photo of an albino SDM calf right, cf. Lars Gjøl Christensen.
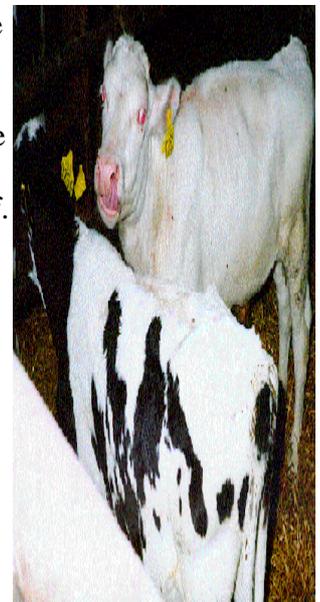
COAT COLOUR, ALBINISM in CATTLE

```
MIA Number :   000202

Possible human homologue (MIM number) : [ 203100 ]

Across-Species Summary :

    Congenital lack of pigmentation in all parts of the body. Due
to a
    non-functional form of the enzyme tyrosinase.

References after 1970
```

Greene, H.J., Leipold, H.W., Gelatt, K.M., Huston, K. (1973). *Complete albinism in beef Shorthorn calves*. Journal of Heredity 64: 189-192.
Weber, W., Lauvergne, J.J., Winzenried, H.U. (1973). *Hereditary albinism in Swiss Simmental cattle* [French]. Schweizer Archiv fur Tierheilkunde 115: 142-144.
Manunta, G., Lai, P., Cancedda, M. (1975). *A contribution to the study of albinism in the Brown Mountain breed*. Zootecnica e Veterinaria. Fecondazione Artificiale 30: 129-135.
Ojo, S.A., Leipold, H.W. (1976). *Ocular albinism in a herd of Nigerian Holstein Friesian cattle*. Zeitschrift fur Tierzuchtung und Zuchtungsbiologie 93: 252-254.
Jayasekera, U., Leipold, H.W. (1981). *Albinism in United States Charolais cattle*. Annales de Genetique et de Selection Animale 13: 213-218.
Foreman, M.E., Lamoreux, M.L., Kwon, B., Womack, J.E. (1994). *Mapping the bovine albino locus*. Journal of Heredity 85: 318-320.

Chapter 2. Hardy-Weinberg law for gene frequency stability in large populations

Genetic terms concerning genes and alleles.
The Labrador Retriever race has the coat colour types yellow and black. Thus considering a gene for yellow coat colour, it points at both a locus and an allele. This corresponds to the early understanding of genes where there was a variant (yellow) corresponding to the wild type (black).

The term 'gene' should not be used when talking about a DNA polymorphism with more alleles in non-coding sequences. Instead the term 'locus' should be used. This terminology also ought to be used in connexion with genetic variants. There is not always a sharp separation between gene and locus, and between gene and allele in practice. So when gene frequencies are mentioned in this chapter, it would include genes in the traditional sense and alleles in non-coding DNA sequences (loci).

2.1 Gene counting method for calculation of gene frequencies

Gene counting method, Co-dominant inheritance, Phenotype = Genotype
Data from a study of albumin types in Danish German Shepherd will be used to explain the principles, see Table 1 section 2.4. When serum is tested on an electrophoresis gel there is a 'fast' and a 'slow' albumin allele designated as the F and S allele, respectively. The investigations have been carried out by K. Christensen et al. 1985, *Hereditas*, 102:219-223.

The following numbers of the 3 albumin genotypes was found in a population.

```
---------
Genotype        SS          SF          FF      Total
Number          36          47          23       106
Frequency      0.34        0.44        0.22     = 1.00
---------
```

The frequency of the SS type is calculated as $36/106 = 0.34$

Figure 2.1. Albumin type is codominantly inherited, which means that both alleles can be seen directly on the gel. The phenotypes, seen on the gel, are an SS, an SF and an FF albumin type. After the electrophoresis of the serum samples the gel is stained with Amidoblack.

The calculation of **gene frequency** for S is based on the gene counting method, every SS individuals have 2 S genes and a SF individual has 1. This is relative to all genes in the population, which are 2*106.

```
The frequency of S is calculated as  p = (2*36 + 47)/(2*106) = 0.56
    do          F                do   q = (2*23 + 47)/(2*106) = 0.44
                                                       ------
                                                        1.00
```

The frequency of the first allele is normally given the symbol p and the second the symbol q. If there are more alleles the symbols are continued alphabetically.

An allele frequency corresponds to a probability, and therefore the sum of alleles (gene frequencies) is 1. For basic statistical formulations, or Samuels, *Statistics for Life Science*

Examples of multiple alleles (more than two alleles)
In section 1.5 the segregation of a mikrosatellite in a swine family was mentioned. There are 3 alleles as can be seen in the gel in Figure 2.2. The allele designations are 209, 199 and 195

Figure 2.2
Segregation of genetic variation (mikrosatellite) in a locus in a swine family detected by gel electrophoresis. The boar Cup carries the alleles 209 and 195 and sow 400 carries the alleles 199 and 195. Most of the animals are heterozygotes each with two bands. The homozygotes have only one band. Gel prepared by Merete Fredholm

The calculations of the gene frequencies by means of the gene counting methods are shown below. For the allele '209' there are 18 heterozygotes and 2 homozygotes (no bands in the lower line). *Count for yourself!* For the allele '199' there are 12 heterozygotes and 0 homozygotes (all 12 have bands in '199' line plus one other band in either '195' or '209'). *Count again!*

```
Frequency of '209' calculated as p = (2*2 + 18)/(2*44) = 0.250
 do              '199'    do        q = (2*0 + 12)/(2*44) = 0.134
 do              '195'    do        r = 1 - p - q         = 0.616
```

The last allele frequency is most easily calculated as a difference, since the sum of the allele frequencies is 1.

The square root method, Dominant inheritance

In a population of Labrador Retrievers the following numbers of the two phenotypes was found:

```
---------
Phenotype       Black           Yellow      Total
Genotype        EE+Ee             ee
Number           182              18          200
Frequency       0.91            0.09        = 1.00
----------
```

Figure 2.3. The Labrador Retriever with the normal colour type black and with segregation of the recessive yellow coat colour type due to alleles in the extension locus.

The black colour type represents both homozygotes and heterozygotes. If two random black dogs are mated, they will occasionally get yellow offspring. This happens when both dogs by random are of the colour genotype Ee.

The gene counting method implies that the phenotype is equal to the genotype, but this is only the case for yellow. For an offspring to become yellow, it should have an e gene from both its sire and dam, each with a probability of q. Therefore, the probability multiplication rule can be applied to calculate the frequency of ee individuals which corresponds to q^2. From the data material it is estimated to be $18/200 = 0.09$

The square root of 0.09 is q = 0.30.

The frequency of the black gene in the population becomes p = 1 - q = 0.7.

2.2 Hardy-Weinberg equilibrium and statistical tests

Definition of H-W equilibrium. In a large population with random mating H-W equilibrium will occur after one generation provided that the same gene frequencies occur in both sexes. Hardy-Weinberg equilibrium implies that gene and genotype frequencies are constant from generation to generation. If disequilibrium occurs, equilibrium will be reestablished after one generation of random mating. The H-W conditions also imply that when the gene frequencies are p and q, the genotype frequencies will be respectively p^2, 2pq and q^2 for the dominant, the heterozygotes and the recessive in a two allele system. This can be inferred by the arguments given for the recessive type under dominant inheritance

Statistical test for H-W equilibrium
Exemplified by data from the albumin types in Danish German Shepherd population, shown in section 2.1. The following observed numbers of the three genotypes (obs), and the calculated expected numbers under Hardy Weinberg equilibrium (exp) with frequencies p(S)=0.56 and q(F)=0.44 was found in the population.

```
---------
Genotype        SS              SF          FF      Total
Number, obs     36              47          23    = 106 = N
Frequency, exp  p²              2pq         q²    = 1,00
Number, exp     p²N             2pqN        q²N   = N
Number, exp     33.2            52.3        20.5   = 106
Deviation       2.8            -5.3         2.5
Chi-square      0.24            0.54        0.31   = 1.09
----------
```

The expected numbers for Hardy Weinberg equilibrium (exp) are calculated using the multiplication rule for probabilities.
For the SS genotype an S gene from a father and an S gene from a mother has to be drawn, the same repeated 106 times. Therefore the probability for an SS genotype is p*p*106. The corresponding arguments

for the other two genotypes lead to the results shown in the table above.

Now a Chi-square test for H-W equilibrium can be calculated as the sum of squared deviations, each divided by the expected number.

The test has 1 degree of freedom as there are three classes, and two free parameters given by the material, p and N, has to be applied to calculate the expected numbers. The last used parameter (q) is not free, as it can be calculated as (1 - p).

By use of the chi-square table (chapter 13), with DF=1 the value 3,84 equals the 5% limits for maintaining the H_0 hypothesis, that the data has H-W proportions. Therefore the found deviation between observed and expected numbers has a probability which is larger than 5%. Conclusion: There is no statistical deviation from H-W equilibrium.

For an applet for calculation of Chi-squares for H-W equilibrium, [click here](#).

2.3 Sex-linked inheritance

One of the preconditions for H-W equilibrium is random mating in the population. In such a population H-W equilibrium always occurs after one generation of random mating, provided that both males and females have the same gene frequency.

By sex-linked inheritance there is not necessarily the same gene frequency in males and females after random mating, since the two sexes get their genes from two different sources.

In mammals the males get all their sex-linked genes from their mothers, whereas the females get half their genes from their sires and half from their dams. The opposite is the case in birds, where the female is the heterogametic sex.

Example, the orange gene in cats: In a population the following number of the three genotypes were found. The orange gene (O) gives yellow coat colour. The genotype Oo gives a mixed colour, which is caused by random inactivation of the X-chromosome in XX individuals. The oo phenotype is non-yellow.

```
---------       ------      females  ---------      ----- males   -----
Genotype     OO       Oo      oo    Total      O       o     Total
Number        3       53     117     173       28     149     177
Frequency   0.02     0.31    0.67  = 1,00      0.16    0.84 = 1,00
----------
```

In mammals the sex-linked gene is shown directly in the males. That is, the genotype frequency equals the gene frequency. Whereas in the females the calculation of the gene frequency is identical with the calculation for autosomal genes, as shown below.

```
Frequency of O calculated as p = (2*3  + 53)/(2*173) = 0.17
   do         o           do    q = (2*117+ 53)/(2*173) = 0.83
                                                         ------
                                                          1.00
```

There are only minor differences between the frequency in the two sexes, as shown. If there is a difference it will be halved in each new generation with random mating, so in practice H-W equilibrium is reestablished after a few generations, if the population deviated from equilibrium in the first place. Phenotypically the heterozygotes are very special, since only one gene is active in each cell. The phenotypes occur because of a random blend of cells with an alternative activated gene.

From the chromosome preparation it can be seen that one of the X-chromosome is inactive (circled), figure 2.4 left. To the right is a heterozygotic female cat with the genotype Oo. It has yellow and black spots caused by alleles in the sex-linked orange locus. The white colour is caused by an autosomal gene for spots.

Figure 2.4
Heterozygote female cat with the genotype Oo in the sex-linked Orange locus. Photo Bodil Andersen, Bogense. It gives a blended coat colour of yellow and black with variating spot sizes. This is caused by random inactivation of one X-chromosome in female mammals. The inactive swine X-chromosome is circled.
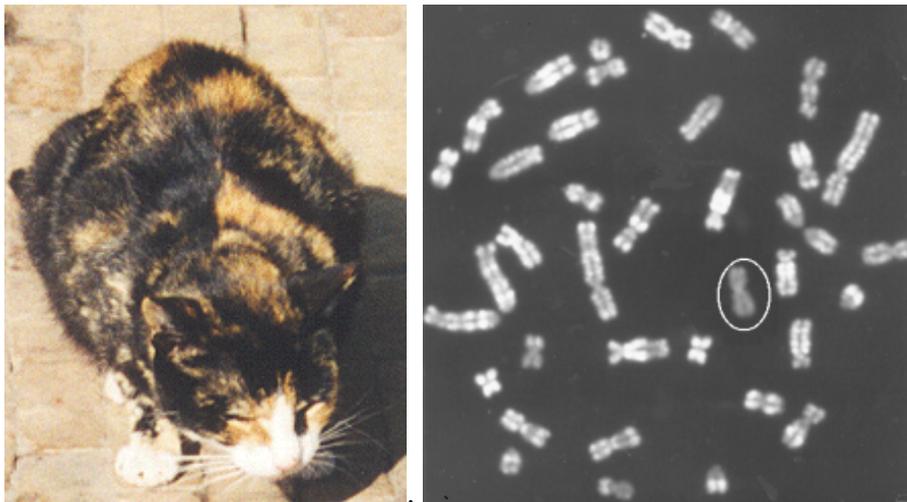
The inactive X-chromosome is detected by means of Acridine orange staining on cells grown for 6.5 hours with BrdU (Brom deoxy-Uridine) in the media. DNA-syntheses which occur after adding BrdU can be seen as weakly stained areas on the chromosomes. There is a circle around the inactive X chromosome. The active genes replicate early in the cell cycle and are therefore found in the white band.

The majority of cases with recessive sex-linked diseases occur in males. Haemophilia in human is one of the best known examples of sex-linked recessive inheritance, the frequency in boys being 100 times larger than the one in girls. This occurs when the gene frequency is 0.01, corresponding to the frequency in boys. Whereas the gene in the girls has to come from both father and mother, each with a probability of 0.01, this corresponds to a frequency in girls of 1 in 10'000.

2.4 Examples of application of gene frequencies

When the gene frequencies in a population are known, the genotype frequencies for each genotype under H-W conditions can be calculated. Correspondingly the mating type frequencies under random mating can be calculated by means of the multiplication rule for probabilities, results shown in the table below. Note that when the two genotypes are different, there has to be multiplied with 2, as the two types can be combined in reverse orders.

Mating type	Frequency		Offspring segregation		
---	---	---	AA	Aa	aa
AA x AA	p^2 x p^2	$= p^4$	1		
AA x Aa	2 x p^2 x 2pq	$= 4p^3$ x q	0.5	0.5	
Aa x Aa	2pq x 2pq	$= 4p^2$ x q^2	0.25	0.5	0.25
AA x aa	2 x p^2 x q^2	$= 2p^2$ x q^2		1	
Aa x aa	2 x 2pq x q^2	$= 4pq^3$		0.5	0.5
aa x aa	q^2 x q^2	$= q^4$			1

Calculation of mating type frequencies is important in relation to control of monogenetic inherited diseases. This topic will be discussed more detailed in chapter 5, where the classical segregation in offspring will be submitted to statistical analysis.
An applet for calculation of mating type frequencies can be seen below

Example of use of the applet is shown below where q=0.005. The 99% of the recessive offspring come from normal parents.

Mating type	Total Freq.	Offspring AA	Offspring Aa	Offspring aa
AAxAA	0.9801495	0.9801495	0	0
AAxAa	0.0197014	0.0098507	0.0098507	0
AaxAa	0.0000990	0.0000247	0.0000495	0.0000247 99.00 %
AAxaa	0.0000495	0	0.0000495	0
Aaxaa	0.0000004	0	0.0000002	0.0000002 0.99 %
aaxaa	0.0000000	0	0	0.0000000 0.00 %
Calculate		sum = 0.9900249	sum = 0.0099500	sum = 0.0000250
Freq. p 0.995				

Gene frequencies also have a large area of application within bred studies. The gene frequencies for closely related breeds have a tendency to lie closer each other than not so closely related breeds.

Table 1. Distribution of albumin types among pure dog breeds and mongrels arranged according to the frequency of the slow allele.

| Breed | Genotypes | | | | Frequency |
	SS	SF	FF	Total	of S
Basset Hound	0	2	30	32	0.03
Beagle	3	14	52	69	0.14
Dachshund	2	8	26	36	0.17
Small Münsterländer	2	3	12	17	0.21
Sheepdog, Shetland	0	6	5	11	0.27
Dalmatian	2	3	10	15	0.30
Collie	2	21	18	41	0.33
Cocker Spaniel	7	24	20	51	0.37
Grey Hound	4	6	8	18	0.39
St. Bernhard	3	8	7	18	0.39
Labrador Retriever	8	10	10	28	0.46
German Shepherd	36	47	23	106	0.56
Old Danish Pointer	5	5	2	12	0.63
Terrier, Tibetanian	10	11	3	24	0.64
Terrier, Airedale	5	12	0	17	0.65
Poodle, Miniature	9	8	2	19	0.68
Newfoundland	25	33	3	61	0.68
Poodle	39	36	6	81	0.70
Boxer	54	14	1	69	0.88
Carelian Bear Dog	15	4	0	19	0.89
Golden Retriever	53	3	1	57	0.96
Great Dane	9	0	0	9	1.00
Rottweiler	10	0	0	10	1.00
Basenji	44	0	0	44	1.00
Other pure breeds	94	57	38	189	0.65
Mongrels	22	41	24	87	0.49
Total	463	376	301	1140	
Genotypic frequencies	0.406	0.330	0.264		
Expected genotypic frequencies	0.326	0.490	0.184		

The calculated gene frequencies were p(S) = 0.571

The above shown table gives the results from a study of the albumin types in Danish dog breeds. As mentioned in section 2.1 a fast and a slow albumin allele occur when serum is tested on an electrophoreses gel. These alleles have the designations F and S. The frequencies of the S allel varies from 0 to 1 in the dog breeds. The investigations have been carried out by K. Christensen et al. 1985; *Hereditas*, 102:219-223.

In the table below are the results from a study of some mikrosatellites in Danish dog breeds, as well as in the red and the blue foxes. Here are results from three loci each with several alleles, being named as the length of the DNA-string being amplified by the primers. The frequencies vary significantly between the dog breeds. The primers which are developed in the dog work equally well in both fox species. This indicates a large degree of genetic homology between dogs and foxes. H[e] gives the information content of the loci. This depends on the percentage of heterozygotes in the population concerned. Genetically the heterozygotic animals are the most informative. In paternity tests or other genetic investigations the homozygotes give very little or no information about which of the parents' alleles an offspring has inherited. The investigations have been carried out by M. Fredholm et al. 1995; *Mamalian Genome*, 6:11-18.

Table 2. Allele sizes and allele frequencies of canine microsatellites.

Microsatellite locus	Core sequence	Allele	Flat-coated retriever[a]	Dachs hund[b]	Arctic fox[c]	Red fox[d]
CPH1	$(TGG)_3$	125			1.00	
	- TAG$(TGG)_{11}$	132				0.30
		135		0.14		
		138				0.70
		141	0.83	0.56		
		144	0.17	0.30		
H[e]			0.21	0.56	0.00	0.40
CPH2	$(AC)_{15}$	95	0.20			
		97	0.36	0.36		
		99	0.20	0.06		
		103		0.06		
		107		0.52		
		109	0.24			
H[e]			0.94	0.47		
CPH3	$(GA)_2TA(GA)_{17}$	154			0.15	0.05
		158			0.10	0.25
		162	0.29		0.45	0.30
		164			0.15	0.40
		166			0.15	
		168	0.09	0.28		
		172	0.02	0.03		
		174		0.02		
		176	0.32	0.28		
		178	0.20	0.20		
		180	0.02	0.14		
		182	0.08	0.05		
H[e]			0.67	0.63	0.80	0.90

2.5 Gamete frequencies under linkage

The term gamete frequency is applied when the alleles from more than one locus are considered. Every new gamete contains an allele from each locus. Consider the two loci, A and B, the genotypes on each locus can occur in H-W proportions, even though this is not the case for the two loci together. This can be of great significance in relation to marker genes (loci). Occasionally a specific allele on the marker loci occurs with the disease allele of another locus. If the two loci are closely linked it could take several generations before equilibrium is reached between the two loci.

The combination of two gene pairs, A and B, each with two alleles are shown below in the classical two-by-two table, where r, s, t, and u are the observed gamete frequencies of the gametes AB, Ab, aB and ab, see Figure 2.5.

```
gene A    gene B       B                b            frequency
          ----------------------------------------------------
   A       |  r=p(A)*p(B)+D    s=p(A)*q(b)-D  |    p(A)
   a       |  t=q(a)*p(B)-D    u=q(a)*q(b)+D  |    q(a)
          ----------------------------------------------------
Frequency  |     p(B)              q(b)       |      1
```

Gene frequencies can be calculated by means of the gene counting method and they correspond to the border distributions. The expected frequency of a gamete is the product of the border distributions which is equal to $p(A) = r + s = p(A)*p(B) + D + p(A)*q(b) - D = p(A)[p(B) + q(b)] = p(A)$. The deviation which occurs between observed and expected numbers is assigned the symbol D (Disequilibrium). Note that the deviations are the same size for all cells (D), but with a negative sign for the repulsion phase gametes, Ab and aB.

A calculation example: The observed number of gametes is AB=21, Ab=49, aB=19, ab=11.

```
gene A/gene B          B                 b              Sum (Freq)
           -------------------------------------------------------------
       A    |    21  (r=0.21)     49  (s=0.49)   |   70  (p(A)=0.7)
       a    |    19  (t=0.19)     11  (u=0.11)   |   30  (q(a)=0.30)
           -------------------------------------------------------------
   Sum (Freq) |    40 (p(B)=0.4)        60  (q(b)=0.6) | 100     1
```

Frequencies are given within brackets: Chi-squared = 9.7** Df=1

$$D= u-q(a)*q(b)=0.11 - 0.3*0.6= -0.07$$

Statistically significant deviation occurs from random combination between the genes A and B corresponding to a disequilibrium equal to -0.07.
The corresponding gametes from the first table can be organized differently as shown below:

```
    gamete    observed frequency    expected frequency    deviation
    -------------------------------------------------------------
     AB              r                  p(A)*p(B)              D
     Ab              s                  p(A)*q(b)             -D
     aB              t                  q(a)*p(B)             -D
     ab              u                  q(a)*q(b)              D
```

In the table below all possible genotypes in a two gene systems are shown. Physical recombination only has significance in the double heterozygotes, in all other genotypes a recombination would not give rise to other types of gametes than those from which they are formed. The corresponding genotype frequencies under random mating are obtained by multiplication of gamete frequencies as shown. The multiplication is done by means of an applet designed to study linkage and linkage disequilibrium.

gamete types	Mother(s)	Freq. gamete AB	Freq. gamete Ab	Freq. gamete aB	Freq. gamete ab
Father(s)	Frequencies	0.1	0.2	0.3	0.4
Freq. gamete AB	0.1	AABB 0.010000	AABb 0.020000	AaBB 0.030000	AaBb 0.040000
Freq. gamete Ab	0.2	AAbB 0.020000	AAbb 0.040000	AabB 0.060000	Aabb 0.080000
Freq. gamete aB	0.3	aABB 0.030000	aABb 0.060000	aaBB 0.090000	aaBb 0.120000
Freq. gamete ab	0.4	aAbB 0.040000	aAbb 0.080000	aabB 0.120000	aabb 0.160000
Recomb. freq.	p(A)= 0.300000	Generation = 1.0	Next generation	Knud Christensei	sept.12.1997
0.5	p(B)= 0.400000	Diseq =-0.020000	Initiate Mendel	Restart	E-mail kc@kvl.dk

Figure 2.5 shows the two types of double heterozygotes. By recombination gametes of the other type is formed and reverse. Therefore there is linkage equilibrium when equal numbers of the two types of double heterozygotes occur. This can also be seen in the formula for D given below.

The expected frequency of a specific gamete corresponds to the frequency where the two genes are inherited independently. In that case the multiplication rule for probabilities can be applied to calculate the expected numbers. The disequilibrium has the symbol D and is defined as shown above (observed - expected gamete frequency). It can also be calculated as half of the difference between the frequencies of double heterozygotes in linkage and repulsions phase:

Figure 2.5. The two homologous chromosomes in linkage and in repulsions phase. When recombination takes place (with a frequency c) the opposite gamete is formed


```
D = u  -  q(a)*q(b), or
D = r*u - t*s    (= [p(AB/ab) - p(Ab/aB)]/2)
```

The maximum disequilibrium exists when all double heterozygotes are in linkage phase (AB/ab) or in repulsion phase (Ab/aB); then it has the value D_{Max}.

If the recombination frequency is c, D will be inflated to $D*(1-c)$ per generation, so in the n'th generation $D_n = D_0*(1-c)^n$, where D_0 is the disequilibrium in the base population, see figure 2.6.

For two independent loci having .5 in recombination frequency 4 to 5 generations will pass before equilibrium is reached. Concerning closely linked loci with less than 5 % recombination, it will take more than 25 generations before equilibrium is reached.

Figure 2.6. Curves of number of generations towards equilibrium for different recombination frequencies.

The release of linkage disequilibrium can produce new genetic variation for quantitative traits. If two loci are linked, both carrying a positive and a negative effect and exist in linkage phase, new variation arises if the positive alleles are recombined to be in linkage phase

Mendel's second law predicts that two loci would segregate independently in F_2 in an analysis cross. This is caused by the fact that all individuals in F_1 are double heterozygotes. Since only the double heterozygotes can recombine, it will take long time before equilibrium arises in a mixed population, where the frequency of the double heterozygotes is only $2(r*u + t*s)$.

In relation to a new mutation complete disequilibrium exists in several future generations, as the mutation arises in only one chromosome. This disequilibrium can be used in relation to application of gene markers to identify the localization of the gene and to identify carriers in related individuals by means of the marker locus.

19

Within a family maximum disequilibrium always occur. This makes it possible to utilize marker genes within families. If an individual is born with a recessive disease and has a specific marker combination, new offspring having the same marker gene combination, would with great probability also carry the disease.

for studying linkage and linkage disequilibrium, click here.

Chapter 3. Deviations from Hardy-Weinberg equilibrium

3.1 Systematic deviations from H-W equilibrium

In a large population with random mating H-W equilibrium occur unless the population is subjected to systematic effects, which can change the gene frequency. The systematic processes can be divided into the categories show in Figure 3.1.

Figure 3.1. Mutation,

Migration and

Selection
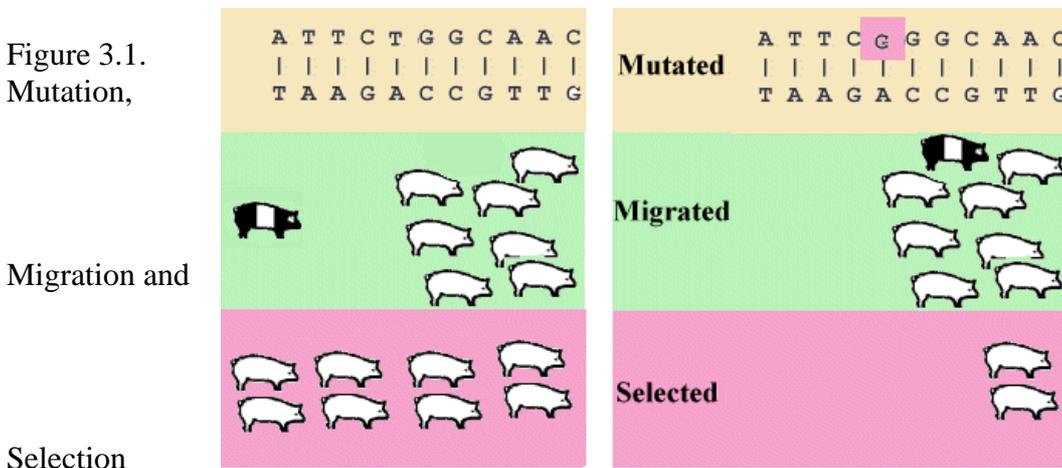
For the proper genes (structure genes) **mutations** generally occur with a frequency of .1 to 1 per million gametes per gene. It occurs 1000 times more often in the number of repetitions in short repeated DNA sequences (mikrosatellites). These sequences are not translated into protein but works as internal spacers between the genes. The sum of mutations in all genes will cause each new individual to carry one or more new mutations. Most new mutations in the structure genes are harmful. Therefore due to low fitness a selection of animals carrying the harmful mutations will occur in every generation. Over a large number of generations a balance between new mutations and selection will be established. For recessive genes the equilibrium gene frequency (q) can be calculated by the following equation:

$$q^2 * s = my,$$

the mutation frequency (my) is equal to the frequency of recessive individuals being removed in each generation ($q^2 * s$). For dominant genes the equilibrium gene frequency (p) can be calculated from the following equation:

$$(1/2)2pq * s = ps = my,$$

the mutation frequency (my) is equal to half of the frequency for heterozygotic individuals, which is removed in each generation (pq*s) and q is approximately 1. For definition of s see section 3.2.

In natural populations **migration** often occurs from the closest neighbourhood. Thereby a smooth transition in the gene frequency will occur between the sub populations. In animal breeding migration corresponds to introduction of new animals. They are normally bought where the best animals are bred.

Natural **selection** corresponds to the term 'survival of the fittest', known from Darwin's theory of evolution. The natural selection will seldom lead to gene fixation, as a population often are subjected to sudden changes in environmental conditions. Therefore it can be an advantage for a population to carry genes which, in the present situation, are not the optimal. Recessive genes on the other hand can be carried in low frequency with only minor loss of vitality (fitness).

The term **genetic load** is used to describe a population with 'lethal genes' or or genes with lower fitness. A population can only carry a certain load which is removed by selection in each generation. In contrast to new mutations, which play a minor role, it is more costly to maintain systems, where overdominance occurs by removal or partly removal of both homozygotic types in each generation.

In the Belgian Blue breed a recessive mutant (double muscles) is found. This breed is shown in Figure 3.2. The gene can only be passed on if a high number of Cesarian sections are accepted as the calves are too big to be born the natural way. In nature most of these calves would die, therefore in nature the gene would only exist in a very low frequency.

Figure 3.2
Bull of the Belgian Blue breed which is double muscled. The term 'genetic load' is visualized.

In the next sections formulas will be given for how the gene frequencies change when a population is exposed to selection pressure on a specific genotype.

3.2 Selection against the recessive

By selecting against the recessive the following table is obtained. Selection (s) against the recessive is relative compared to the dominant types. The proportion selected of a given genotype is given the symbol s, which do not reproduce in every generation. Therefore, the fitness is equal to 1-s.

```
Table formulating selection:
---------
Genotype          EE              EE              ee        Total
Frequency         p²              2pq             q²        = 1,00
Fitness           1               1               1-s
Proportion        p²              2pq             q²(1-s)   = 1-sq²
after selection
---------
```

After selection the gene frequency is calculated by the gene counting method. The new gene frequency q' is estimated as the heterozygotes plus 2 times the surviving recessive relative to 2 times the proportion of all surviving genes after selection, which is equal to $1-sq^2$.

$$q' = (2pq + 2*q^2(1-s))/(2(1-sq^2))$$

The frequency q' represents the genes which survive and therefore corresponds to the gene frequency in the next generation before selection. The formula can be applied repeatedly generation after generation. In the right side of the formula q' is calculated in the preceding generation and so forth.

Figure 3.3 shows such an application. By strong selection (s=1) the gene frequencies change very rapidly at high gene frequencies. If the gene frequency in contrast is low, the selection will hardly affect the frequency, the population has parked the gene in waiting position so to speak. This causes some problems if it concerns a recessive disease gene in an animal population. Selection can not really solve this problem. Therefore there is great interest in finding a DNA test method to diagnose the heterozygotic carrier. From Figure 3.3 it also becomes evident that by weak selection pressure the changes in the gene frequency are always very slow.

Figure 3.3. Effect of selection against the recessive and changes in gene frequency through generations. At a low gene frequency the changes are very slow.
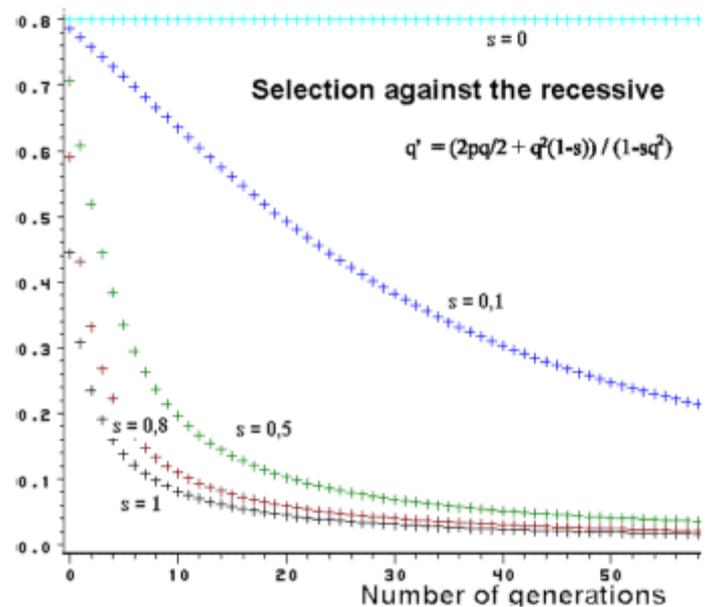
For s=1 the formula for gene frequency changes can be expanded to n generations, when q_0 is the gene frequency in the initial population:

$$q_n = q_0/(1+n*q_0)$$

From this formula n can be isolated and it can be calculated how many generations is needed to obtain a given change in the gene frequency q.

$$n = 1/q_n - 1/q_0$$

Example: When s=1 the gene frequency change from 0.01 to 0.005 would take

$$n = 1/0.005 - 1/0.01 = 200 - 100 = 100$$

It takes 100 generations to change the frequency of a lethal gene from 1 % to 5 per thousand.

For an applet, as seen below, for calculating gene frequency changes for different fitness combinations, click here The Calculations show that the gene frequency change from .5 to .42 after one generation of selection when the fitness of the recessive is equal to .5.

23

Genotypes	AA	Aa	aa
Expected before sel	0.250000	0.500000	0.250000
Fitness Genera=1.0	1.0	1.0	0.5
Expected after selec	0.285714	0.571428	0.142857
p after sel.	q after sel.	Genetic load = 0.125	Calculate !
0.571428	0.428571	delta q =-0.071428	+ 1 generation !

DNA test and culling of heterozygotic males

It takes a very long time to eliminate a recessive disease gene from a population, when there has only been selected against the recessive. In many populations relatively fewer males than females exist. Therefore, it is often enough to DNA test the males. As the males contribute to half of the genes in the new generation, the gene frequency of the disease gene would be halved in every new generation, if only free males are used for breeding. At the same time it is attained that no segregation of recessively affected individuals will occur.

Example: The recessive inherited disease BLAD (Bovine Leukocyte Adhesion Deficiency) occurs in HF dairy breeds. In 1992 the gene frequency in the Danish female population was about 15% of the BLAD gene. By systematically 'BLAD' DNA testing all new bulls the frequency in year 2000 is down to less than 7%. By this procedure the frequency of the BLAD gene is halved in every new generation.

3.3 Selection for heterozygotes

By selecting for heterozygotes (overdominance) the following table is obtained. Selection against the recessive (s_2) and the dominant (s_1) is relative in comparison to the heterozygote types.

```
Table formulating selection for heterozygotes:
---------
Genotype        EE              Ee              ee        Total
Frequency       p²              2pq             q²        = 1,00
Fitness         1-s₁            1               1-s₂
Proportion      p²(1-s₁)        2pq             q²(1-s₂)  = 1-p²s₁ - q²s₂
after selection
---------
```

After selection the gene frequency is calculated by the gene counting method as shown in section 3.2.

$$q' = (2q^2*(1-s_2) + 2pq)/(2*(1-p^2s_1 - q^2s_2))$$

The frequency q' represents the genes that survive and therefore corresponds to the gene frequency in the next generation before selection. In this case, concerning overdominance, selection will not end by fixation of one of the alleles, instead an equilibrium with constant gene frequencies will occur.

The equilibrium frequency is called q (hat), and equilibrium is reached when no change occurs from one generation to the next, i.e. delta q = 0

$$\text{delta } q = q' - q = pq(ps_1-qs_2)/(1-p^2s_1 - q^2s_2) = 0$$
$$ps_1-qs_2 = 0$$

which solved with respect to q gives the equilibrium frequency

$$q(hat) = s_1 / (s_1 + s_2) \text{ or}$$
$$p(hat) = s_2 / (s_1 + s_2)$$

Delta q equals the change in the gene frequency from one generation to the next. When the gene frequency is larger than the equilibrium gene frequency q (hat) delta q becomes negative, and when gene frequency is lower delta q is positive. Therefore selection for heterozygotes is a neverending selection. Thus the population should carry a large genetic load, it is costly to maintain that type of polymorphism.

24

Overdominance is best utilized in animal breeding by producing crossbreeds, in which all individuals can be heterozygotes.

Fitness by overdominance is visualized in Figure 3.4, here the two homozygote types have a fitness level which is lower than the heterozygotes, whose fitness level is 1.

Figure 3.4.
Fitness conditions by overdominance.
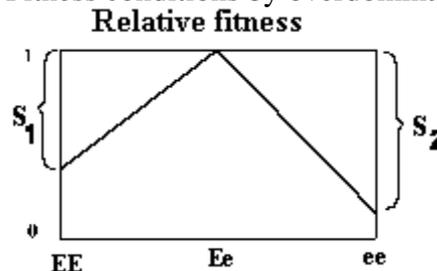
Example: The classical example from human genetics of overdominance is the occurrence of the Mendelian inherited recessive sickle cell anaemia with a frequency of around 5% corresponding to $q = 0.22$ in malaria areas. Individuals which are heterozygotic for the sickle cell anaemia are resistant to malaria, this gives them a greater chance of survival than normal individuals. Individuals with anaemia have low chances of survival, $s_2=1$.
What is the fitness level for the normal homozygotes in comparison to the heterozygotes?

```
Equilibrium occurs at    p(hat) = s₂ / (s₁ + s₂) = 1 - q =1 - 0.22

which gives             s₁ = (s₂ /(1 - q)) - s₂ = 0.285
```

Fitness in normal individuals in a malaria area is $1 - 0.285 = 71{,}5\ \%$ in comparison to the heterozygotes. The 'genetic load' of the population is $p^2s_1 + q^2s_2 = 0.22$, which means that 22% of a generation succumb to maintain the equilibrium, either because of the anaemia or the malaria.

Applet for calculation of gene frequencies for different fitness combinations, click here

3.4 Selection against heterozygotes

The table, which was used in the selection for heterozygotes, can also be used in selection against the heterozygotes. Selection (s) of the recessive and the dominant type now has a negative sign and again it is relative to the heterozygote type, see Figure 3.5.

```
 Table formulating selection for heterozygotes, s₁ and s₂ is now negative:
---------
Genotype           EE                Ee              ee         Total
Frequency          p²               2pq              q²         = 1,00
Fitness            1-s₁              1               1-s₂
Proportion         p²(1-s₁)         2pq              q²(1-s₂)    = 1-p²s₁ - q²s₂
after selection
---------
```

After selection the gene frequency is calculated by means of the gene counting method as shown in section 3.2.

$$q' = (2q^2*(1-s_2) + 2pq)/(2*(1-p^2s_1 - q^2s_2))$$

The frequency q' represents the genes which survive and therefore corresponds to the gene frequency in the next generation before selection. The equilibrium frequency is q (hat), and equilibrium occurs when there is no change from one generation to the next, i.e. delta q = 0.

$$delta\ q = pq(ps_1-qs_2)/(1-p^2s_1 - q^2s_2) = 0$$

$$q(hat) = s_1 / (s_1 + s_2)$$

Figure 3.5

Figure 3.5 shows the fitness level when selection against the heterozygotes has been made. The two homozygote types' fitness level is higher than the heterozygotes, whose fitness is 1.

Fitness condition by selection against the heterozygotes.

Relative fitness
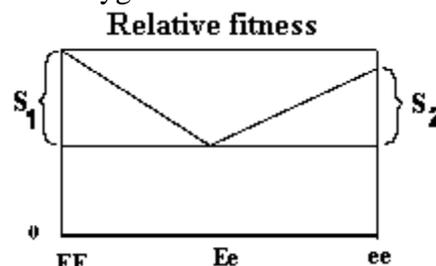
Delta q is the change in gene frequency from one generation to the next. When the gene frequency is larger than the equilibrium gene frequency q (hat), delta q is positive, and when the gene frequency is lower delta q is negative. Therefore selection against the heterozygotes leads to an unstable situation. It leads to fixation of the allele whose frequency is higher than that of the equilibrium. In some cases the heterozygotes can be selected against in one life period and selected for in another, and still equilibrium occurs.

An example of selection against the heterozygotes: When chromosome polymorphism occurs in the Blue foxes, as shown in section 10.3, the heterozygotes individuals have a lower reproduction rate than the homozygotes individuals. But the polymorphism is still found in many natural populations?

3.5 Random deviations from Hardy-Weinberg equilibrium

Problems can occur with the H-W equilibrium in small populations, as a gene frequency can change by chance.
In the smallest possible population (1 male and 1 female), only 4 genes can be carried on to the next generation. In a two allele system with p=0.5, this corresponds to the 5 possible gene frequencies in the next generation, as shown below, when both parents have been heterozygotes (Aa):

```
AA – AA   q=0
AA – Aa   q=0.25
AA – aa   q=0.5
Aa – Aa   q=0.5
Aa – aa   q=0.75
aa – aa   q=1
```

If the gene frequency is 0 or 1, the gene is either fixed or lost. If 4 genes are drawn from a base population with a gene frequency of p =0.5, the chance of loss or fixation (all four genes are a or A) is $0.5^4 = 0.0625$.

Figure 3.6
Standard deviations from a gene frequency depending on population size.

The variance of the gene frequency corresponds to a binominal variance. This can be used in evaluating whether a population is small with random changes in gene frequencies or large with stable gene frequencies.
An applet to simulate random changes in gene frequency depending on population size is found here

$$\sigma_p = \sqrt{\frac{p*q}{2N}}$$

The binominal variance is equal to (p*q)/(2*N), 2*N being equal to the number of genes transmitted to the next generation.
Figure 3.6 shows the standard deviations of the gene frequency (p) when the base population has an average gene frequency of 0.5. The new generation consists of N individuals with 2N genes. The standard deviation has significance for the gene frequency when going from one generation to the next. The random process is important, when only a limited number of animals can carry on the genes. In Figure 3.6 it is clear that in a large population the standard deviation of

the gene frequency is small. Which means that the H-W law is true in the large population, with constant gene frequencies from generation to generation

In cases of extreme low or high gene frequencies the number of individuals has even more significance if a gene is lost or fixed. In the large population even genes with very low frequencies would remain.

Rule of thumbs: It can be derived from Figure 3.6, that large populations are counted in hundreds not in tens.

The size of the population is the problem when considering endangered species. If a specie comes under 100 breeding individuals its chance to survive is when man helps to increase the effective population size, see chapter 9.3.

In the small population, if it survives, fast changes will occur away from the base population. Darwin described the phenomenon as 'evolution by isolation'. He based his theory on observations from the many isolated islands in the Pacific ocean. Here both the animals and plants differed strongly from the one on the continents.

3.6 Effective population size

Every new generation inherits half its genes from the sires and half from the dams. This is very important in, for instance, cattle breeding, where the possibility of artificial insemination is used. The number of bulls can here be very low compared to the number of cows.

In case of uneven numbers of the two sexes the effective population size (Ne) can be calculated according to this formula:

$$4/N_e = 1/N_{sires} + 1/N_{dams} \text{ or solved for } N_e$$

$$N_e = (4N_{sires} * N_{dams})/(N_{sires} + N_{dams})$$

Example 1) 10 sires and 10 dams

$$4/N_e = 1/10 + 1/10 \quad \text{correspond to } N_e = 20$$

Example 2) 1 sire and 10 dams

$$4/N_e = 1/1 + 1/10 \quad \text{correspond to } N_e = 3,7$$

Example 3) 100 sires and 100000 dams

$$4/N_e = 1/100 + 0 \quad \text{correspond to } N_e = 400$$

The variance of the gene frequency in next generation corresponds to what is calculated by use of Ne, see section 3.5. If the number of males and females are the same (example 1), Ne equals the sum. If the number of females is infinite compared to the number of males, Ne equals 4 x the number of males. The sex which has the lowest number of individuals determines the effective size of the population.

The effective population size and increase in inbreeding.
The effective size of the population is important in relation to accumulation of inbreeding in a population. In populations with few animals all the animals will be closely related to each other within a few generations. Inbreeding occurs when the parents of an individual are related, see chapter 4. Inbreeding leads to several negative effects, which are directly proportional to the coefficient of inbreeding. According to the formula the increase in inbreeding is inversely proportional to the effective size of the population:

delta F = 1/(2*Ne)

In a population where Ne = 20 the increase in inbreeding is 2.5% per generation.

Chapter 4. Relationship and inbreeding

4.1 Relationship and inbreeding, definition

As introduction to this chapter should be given some general remarks. Inbreeding is often a negatively charged word, which generally is used to describe situations where no new blood is introduced. As for instance a scientific institute can get the attribute inbred. The expectation is also lower to inbred animals. Relationship is synonymous with 'being in family with'. Relationship is normally a positively charged word as for instance. I'm in family with this person.

Inbreeding occurs in offspring of individuals that are related. Related individuals have common ancestors in the pedigree. Common ancestors can be parents, grand parents etc. These common ancestors are not necessarily from the same generation in relation to the individuals in question. Two individuals are for instance related if the grand father of one of them is the father of the other. In Figure 1 is shown the consequences of identical homozygocity in offspring after full sib mating

Figure 4.1. Identical homozygocity in offspring after full sib mating.
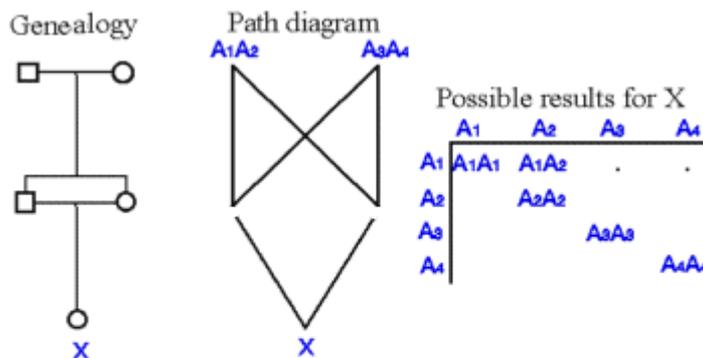
The degree of inbreeding is defined as **the probability that identical homozygocity occurs in a locus**. The identity is not an allele type, but it should be inherited from the same allele in the common ancestor. The degree of inbreeding is given the symbol F and can vary between 0 and 1. The degree of inbreeding in outbred individuals is 0. This is valid even though the parents of the individuals have common ancestors more than 10 generations back. The

degree of inbreeding after full sib mating is 0.25, which can be derived by means of the example in Figure 4.1. **The degree of additive relationship (a)** between two individuals is defined as **two times the probability that two randomly chosen alleles in a locus are identical because of decent from the common ancestor(s)**, or the proportion of identical alleles from the common ancestor(s) in the two related individuals. The alleles are 'identical by decent'. Inbreeding in an individual and the additive relationship between the two parents is directly related. **The degree of inbreeding in an individual constitutes half of the degree of relationship between the two parents.** The degree of relationship and inbreeding are also called **the coefficient of relationship and inbreeding.**

A graph of inbreeding on chromosome level is shown here

4.2 Relationship and inbreeding, calculation examples and formulas

Related individuals can be put into a genealogical diagram. When calculating the coefficient of relationship and inbreeding it is easier to trace the individuals in the pedigree back to the common ancestors using a path diagram. Only individuals of significance for the inbreeding are referred to in the diagram. Which means that individuals that do not lead back to the common ancestors are excluded, see example in Figure 4.2.

The coefficient of relationships (a) is calculated by tracing all possible relations between the two parents through the common ancestors. Note the dotted lines in the figure. Also note that every single animal is being noted throughout the routes. Then the number of generations in each route is counted. The additive relationship can now be calculated as the sum of 1/2 in the power of n (n = number of generations) as shown in Figure 4.2. F_A in the formula, Figure 4.3, points to the coefficient of inbreeding on a given common ancestor, in this case A and B, both having F=0. Therefore the multiplication factor is 1 in both cases, and thus not shown in the calculations. The coefficient of inbreeding is 1/16, which is half of the parents' coefficient of relationship.

Here n designates the number of generations between the parents X and Y, through a common ancestor, and F_A symbolizes the coefficient of inbreeding for the common ancestor belonging to a given route.

Figure 4.2. Example: Calculating relationship and inbreeding after first cousin mating.

E and F are related and the common ancestors for E and F are A and B
A and B are not inbred

Possible routes between E and F over common ancestor

E - C - A - D - F   n = 4 generations
E - C - B - D - F   n = 4 generations

$$a_{EF} = \sum (1/2)^n (1 + F_A) = (1/2)^4 + (1/2)^4 = 1/8$$

$$F_X = \frac{a_{EF}}{2} = 1/16$$
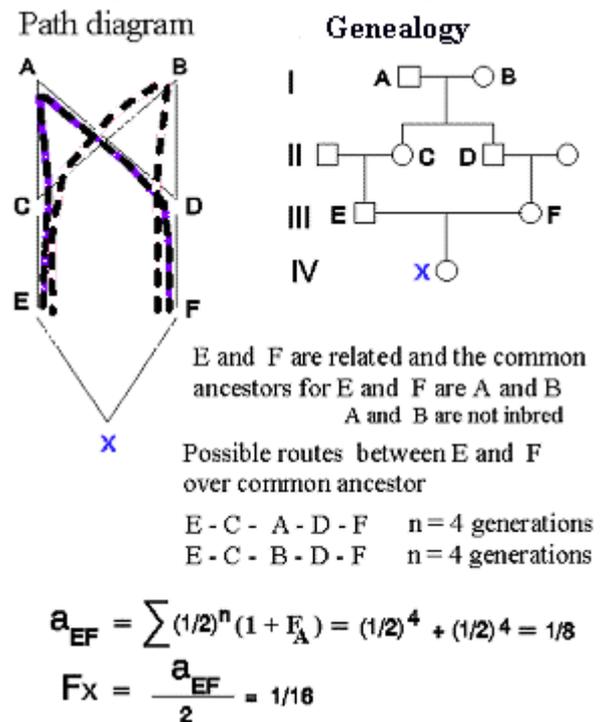
Figure 4.3. Formula for calculation of the coefficient of additive relationship between two animals, X and Y.

$$a_{XY} = \sum (1/2)^n (1 + F_A) \qquad F_Z = \frac{a_{XY}}{2}$$

4.3 Simple forms of close inbreeding

Related individuals can be used as sources of information when estimating the breeding value of an animal, see chapter 7. Therefore it is important to have a clear understanding of which types of relatives occur in large numbers and which at the same time are closely related. Figure 4.4 shows the most common forms of close relationship.

Recapitulation of the calculation methods for the coefficient of relationship: Count the number (n) of generations between the parents A and B through the common ancestors. How many different routes occur? The coefficient of relationship is the sum of $0.5^n(1+F_A)$ for all possible routes.

Example: The coefficient of relationship between full sibs is $0.5^2 + 0.5^2 = 0.50$, from the routes back to each parent.

Figure 4.4. Relationship and inbreeding after the mating of closely related individuals.

The coefficient of inbreeding in the offspring equals half of the coefficient of relationship between the two parents. The figure also shows that parent-offspring have the same relationship as full sibs. The same is true for half sibs, and an individual and its grandparent.

Full and half sibs can be bred in large numbers within domestic animals. New methods for the cloning of oocytes make it possible to produce several identical twins.

When breeding laboratory animals, for instance rats and mice, it is common to reproduce the animals by continued full sib mating. Most of the laboratory strains are 100 % inbred, and the inbreeding is maintained by continued full sib mating. For an outbred population 20 generations of continued full sib mating is necessary, before the coefficient of inbreeding is above 99%. Most of our large domestic animals have difficulties in reproducing if the inbreeding coefficient is above 50%, as inbreeding lowers the fitness, as will be shown in the next section. In chapter 9 more details of the effects of inbreeding and how to control it will be given.

4.4 Segregation of the recessive by inbreeding

The frequency of the homozygotic recessive individuals in the population will increase by inbreeding, this is particularly true when the gene frequency is low.

Example: In a population the gene frequency of a recessive gene is 0.01 corresponding to a frequency of homozygote recessive individuals at 0.0001 or 1 in 10,000. If full sib mating forms the new generation, only, the inbreeding coefficient is 25% in all individuals in the next generation.
In this population the frequency of homozygote recessive individuals can be calculated as follows: It is
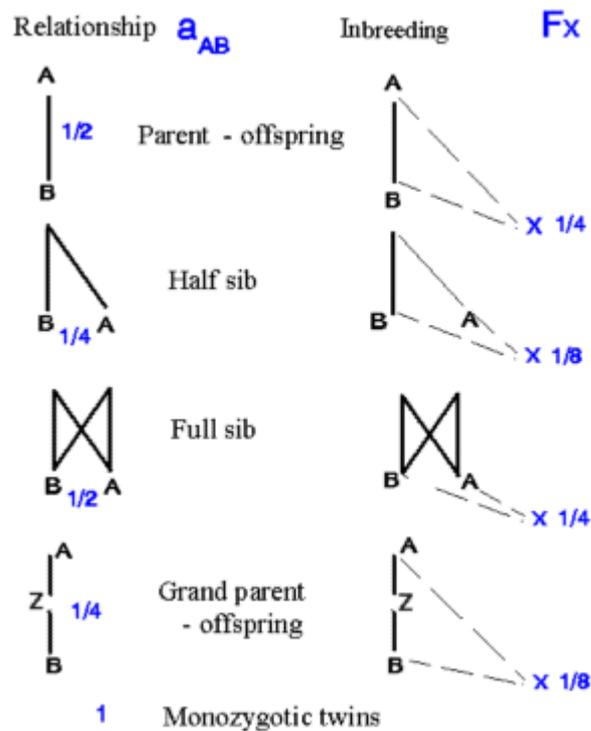
proportional to the number of heterozygotes individuals in the parents' generation ($2*0,01*0,99 = 0,0198$) and their chance that segregation of the recessive occurs (1/16) by full sib mating (Figure 4.1), this segregation should equal one of the genes, the disease gene. Either the grandfather or the grandmother can carry the disease gene. Thus, the joint probability should be multiplied by two.

Therefore the joint probability of breeding homozygote recessive individuals is $2*0,0198/16 = 0,0025$, which means that it has increased 25 times compared to the outbred base population. (In these calculations the case where other combinations of herozygosity and homozygosity of the grandparents are not taken into account, so the result deviate slightly when compared to the general formulas shown below). If the formulas given below are used in our example, the result is $q^2 + pqF = 0.0001 + 0.99*0.01*0.25 = 0.0026$.

Correspondingly, the joint number of homozygote dominants will be increased by the same proportion as the recessive at the expense of the heterozygotic individuals. During inbreeding the proportions would change in comparison to the expected frequency under Hardy Weinberg equilibrium with the following expectations, exp:

```
---------
Genotype        AA              Aa              aa
Frequency, exp  p²              2pq             q²
                +pqF            -2pqF           +pqF
```

By means of the expected frequencies the degree of inbreeding can be calculated in sub populations;

$$\text{which gives} \qquad F = (H_0 - H_n)/H_0 \ ,$$

where H_0 and H_n are the genotype frequencies for heterozygotes in respectively the base population and generation n of the subdivided populations.

Applied on the Albumin example of dogs, section 2.4, the average inbreeding within dog breeds is $F = (0.490 - 0.330)/0.490 = 0.33$.

4.5 Calculation of inbreeding and relationship, the tabular method

Calculation of inbreeding and relationship can be done by means of the tabular method. The big advantage of the tabular method is that a graph of the pedigree is not needed. Simultaneously there is only need to concentrate on two generations at one time, as the method always moves from the older to the younger animals. Therefore it is possible to build on the inbreeding and relationship of the preceding generation, as each level only points back to the parents.

Only two simple formulas are applied in which the relationship and inbreeding have the symbols a and F, as shown in the following two points:

1) Inbreeding in an animal (X) is equal to half the relationship between its parents, A and B, and the additive relationship with oneself is
$a_{XX} = 1 + F_X$
2)The additive relationship between two animals, X and Y, are equal to half of the relationship between the oldest (X) and the parents, A and B, of the youngest (Y).
$a_{XY} = (a_{XA} + a_{XB})/2$

The construction of the relationship matrix are shown in the following points.

1) Put the number of all animals in a line (the oldest first)
2) Put the number of each individual's parents above the number of the individual.
3) Copy step 1) in a vertical column at the left side of the paper
4) Write a row of 1's diagonally

5) Calculate the relationship between animal 1 and 2, 3, 4 etc. to n
6) Copy the first row in the first column
7) Calculate the relationship between animal 2 and 3, 4 etc. to n
8) Copy the second row in the second column
9) etc. - - half of the parents relationship is added to the diagonal element, if relevant step 9 shall be carried out before step 7

Example of application of the tabular method
Using the following set of data.

```
Animal  Sire  Dam
    1     0    0
    2     1    0
    3     1    0
    4     1    0
    5     3    2
    6     3    4
    7     5    6
```

For hand calculation the steps from 1 to 4 is shown first. (-) means an unknown parent

```
Parents          - -    1 -    1 -    1 -    3-2    3-4    5-6
Animal           1      2      3      4      5      6      7
               -----------------------------------------
    1           1
    2                  1
    3                         1
    4                                1
    5                                       1
    6                                              1
    7                                                     1
```

The next steps (5 and 6) is to fill out the first row. The first value is the value for animal number 2. Look at the column of the parents, which shows 1 and - (unknown). The value in column 1 is 1, and unknown is equal to 0, which equals $(1 +0)/2 = 1/2$ which is put into column 2. The same result is obtained for animal 3 and 4. For animal 5 the parents are 2 and 3, in their corresponding columns is 1/2 and 1/2, which added and divided by 2 gives 1/2. The results are the same for animal 6 and 7. Now row one is copied into column one as seen below.

```
Parents          - -    1 -    1 -    1 -    3-2    3-4    5-6
Animal           1      2      3      4      5      6      7
               -----------------------------------------
    1            1     1/2    1/2    1/2    1/2    1/2    1/2
    2           1/2     1
    3           1/2            1
    4           1/2                   1
    5           1/2                          1
    6           1/2                                 1
    7           1/2                                        1
```

Then continue to row two (step 7 and 8). Animal 3 and 4 have 1 and unknown as parents. In column one second row is 1/2, which equals $(1/2+0)/2 = 1/4$ for animal 3 and 4. For animal 5, which have the parents 2 and 3, we have 1 and 1/4 in the two corresponding columns, this gives $(1+1/4)/2 = 5/8$. Animal 3 and 4 are parents of animal 6, their average relationship to 2 is 1/4, which is put into the diagram. Animal 7 has the

parents 5 and 6, their average relationship to 2 is (5/8+1/4)/2 = 7/16, which is put in. The new results are copied to the corresponding places in column 2.

Parents	- -	1 -	1 -	1 -	3-2	3-4	5-6
Animal	1	2	3	4	5	6	7
1	1	1/2	1/2	1/2	1/2	1/2	1/2
2	1/2	1	1/4	1/4	5/8	1/4	7/16
3	1/2	1/4	1				
4	1/2	1/4		1			
5	1/2	5/8			1		
6	1/2	1/4				1	
7	1/2	7/16					1

Now we can continue filling in the table by repeting steps 7 and 8. Remember (step 9) that half of the relationship between the parents should be added to the diagonal element. This is not relevant until animal 5, where the parents, 2 and 3, have a relationship of 1/4, which can be read in row 2 column 3. Remember, if relevant step 9 should be carried out before step 7.

Parents	- -	1 -	1 -	1 -	3-2	3-4	5-6
Animal	1	2	3	4	5	6	7
1	1	1/2	1/2	1/2	1/2	1/2	1/2
2	1/2	1	**1/4**	1/4	5/8	1/4	7/16
3	1/2	1/4	1	1/4	5/8	5/8	5/8
4	1/2	1/4	1/4	1	1/4	5/8	7/16
5	1/2	5/8	5/8	1/4	1+1/8		
6	1/2	1/4	5/8	5/8		1	
7	1/2	7/16	5/8	7/16			1

The final table:

Parents	- -	1 -	1 -	1 -	3-2	3-4	5-6
Animal	1	2	3	4	5	6	7
1	1	1/2	1/2	1/2	1/2	1/2	1/2
2	1/2	1	1/4	1/4	5/8	1/4	7/16
3	1/2	1/4	1	1/4	5/8	5/8	5/8
4	1/2	1/4	1/4	1	1/4	5/8	7/16
5	1/2	5/8	5/8	1/4	1+1/8	7/16	25/32
6	1/2	1/4	5/8	5/8	7/16	1+1/8	25/32
7	1/2	7/16	5/8	7/16	25/32	25/32	1+7/32

The tabular method is simple to computerize, so this method should be used in large animal populations.

The path graph is best used in relation to the genealogy, where good overview is important. In such cases the individual animal can play an important role.

The same results calculated by applet. Input in lower part of the window, output in the upper part.


```
rel =
 1.000 0.500 0.500 0.500 0.500 0.500 0.500
 0.500 1.000 0.250 0.250 0.625 0.250 0.437
 0.500 0.250 1.000 0.250 0.625 0.625 0.625
 0.500 0.250 0.250 1.000 0.250 0.625 0.437
 0.500 0.625 0.625 0.250 1.125 0.437 0.781
 0.500 0.250 0.625 0.625 0.437 1.125 0.781
 0.500 0.437 0.625 0.437 0.781 0.781 1.218

a =
 1  0  0
 2  1  0
 3  1  0
 4  1  0
 5  2  3
 6  4  3
 7  5  6
rel = :tabular(a)

3   clear output   list vars   clear vars   evaluate
```

It is easier to use a program when carrying out the calculation, click here for an applet.
A DOS program, which can be used to analyses entire populations, can be seen here. To download the program, click here, and for an error list file click here. When your browser wants to save, click save and save the files 'inbred.exe' and fl90.eer. The file 'inbred.exe', which is a dos program, can now be executed in DOS. The input file should be generated as shown in the printed example.

Chapter 5. Test of simple genetic hypotheses, experimental or field data

5.1 Genealogy and formulating genetic hypotheses

When new traits or diseases are recognized it is important to formulate genetic hypotheses as an alternative to other exogene causes. Exogene factors can be malnutrition, housing, microbiological infection or a combination of exogene and endogene (genetic) factors.

When a genetically caused disease is likely to occur, some characteristics have to be present. As for instance a family wise occurrence, and the fact that only some members of the family are affected. The typical Mendelian segregation ratios are 1:1, 1:3 and 1:7. In these context traits with intermediary inheritance is not included, as these types of inheritance are rarely seen in newly discovered diseases.

An alternative to Mendelian segregation are the so-called threshold diseases, where the sum of environmental and genetic factors triggers the disease. Threshold diseases with low frequency will exhibit family wise occurrence, which is very similar to Mendelian segregation, see section 8.3, where the inheritance of threshold traits is described.
The schematic overview of the different types of inheritance given in section 1.2 should furthermore be taken into consideration.

The simple Mendelian inheritance can occur with four different forms of inheritance. The traits can be inherited as dominant or recessive, compared to the normal type, and can exhibit autosomal or sex-linked inheritance. On the following pages the characteristics of the four simple forms of inheritance will be demonstrated. Note that the mentioned Mendelian inheritance indications can only be applied if the population frequency is low. This is always the case in newly discovered genetically caused diseases within one or a few families.

Concerning diseases with high population frequency, as for instance breast cancer in the human population, there will often be a significant amount of heterogeneity. This disease has a genetic background, which can be caused by segregation of genes in more independent loci. When large heterogeneity occurs the segregation analysis can not be carried out between families, but only within one big family. Finally there is a problem with **Phenocopies**, which is an identical disease caused by non genetic causes. Phenocopies also occur in breast cancer in humans.

Figure 5.1. Symbols used in genealogy.

The different symbols which are applied in genealogy are shown in Figure 5.1.

The proband individual is the individual that caused the initiation of the investigation. This individual or family should normally be excluded in later statistical analyses. A combination of path diagram and genealogical diagram can also be utilized along with the genealogical diagram.

5.2 Autosomal recessive inheritance

Factors which indicate autosomal recessive inheritance.

- 1. The individuals which are affected are often inbred

- 2. Affected individuals do not necessarily occurring in every generation
- 3. All offspring with both parents affected are also affected
- 4. The population frequency in males and females are the same
- 5. The segregation is 1:3 in offspring of normal parents
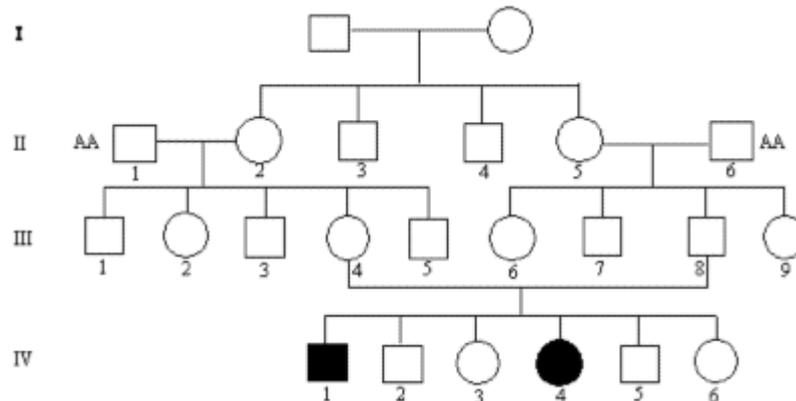
Figure 5.2.

Figure 5.2 shows an example of autosomal recessive inheritance.
Genes, which coded for enzymes, are typical example of recessive inheritance, since one normal gene is often sufficient for the individual to function normally. Increased transcription rate can compensate for the one defect code. See also the example

5.3 Autosomal dominant inheritance

Factors indicating autosomal dominant inheritance.

- 1. Affected individuals occur in every generation
- 2. Affected offspring must have at least one affected parent
- 3. Normal offspring of affected parents get normal offspring
- 4. The population frequency is the same in males as in females
- 5. The segregation ratio is 1:1 among offspring of one normal and one affected parent
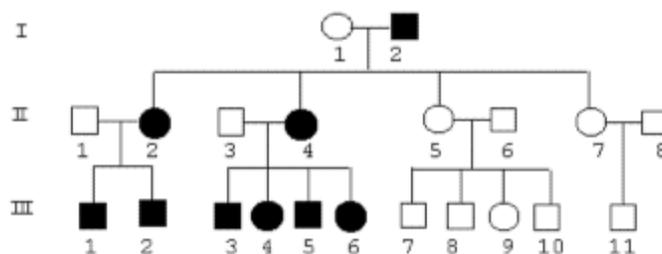
Figure 5.3.

For examples of autosomal dominant inheritance, see Figure 5.3. Typical examples of autosomal dominant inheritance are genes coding for membrane proteins, building blocks in tissues etc.

5.4 Sex-linked recessive inheritance

Sex-linked inheritance works differently in mammals than in birds. In mammals the sex-chromosomes are called XX and XY in female and male, respectively. In the birds the female is the hetero gametic, WZ, and the male is homo gametic, ZZ. The sex of the offspring is determined by the hetero gametic sex, in birds by the female and in mammals by the male.

Factors indicating sex-linked recessive inheritance.

- 1. Affected females are often inbred
- 2. Affected individuals do not necessarily occur in every generation
- 3. Offspring of two affected parents are also affected
- 4. In offspring of normal parents, only males can be affected
- 5. The population frequency in males is higher than in females
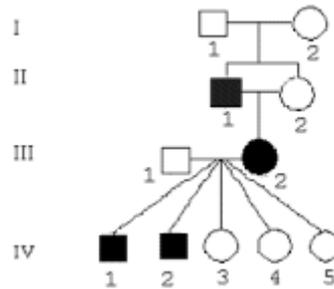
Figure 5.4.

See the example of sex-linked recessive inheritance in Figure 5.4.

5.5 Sex-linked dominant inheritance

Factors indicating sex-linked dominant inheritance.

- 1. All female offspring of an affected father will be affected
- 2. All male offspring of affected fathers will be normal
- 3. Normal offspring of affected parents produce normal offspring
- 4. The population frequency in males is lower than in females
- 5. The segregation ratio is 1:1 in offspring of a normal father and an affected mother

In Figur 5.5 is shown X-ray pictures of Swedish-Danish farm dog affected with spondylo-epifysal dysplasia. The pictures is taken at the X-ray clinic at this university by J. Arnbjerg. The condition is likely to have a sex linked recessive inheritance.
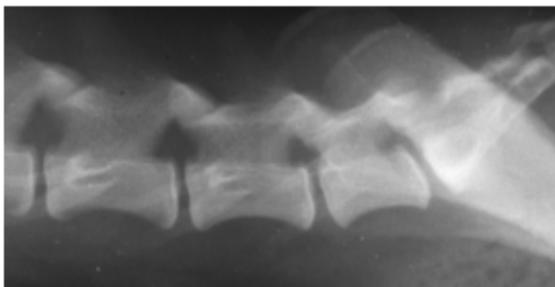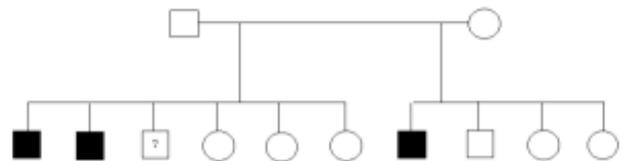
Figur 5.5.   Normal   Affected

5.6 Test mating, statistical tests

When a smaller material is pointing at Mendelian segregation, there would often be alternative hypotheses, which cannot be excluded. The segregation can, as mentioned earlier, also be caused by simulated genetics (phenocopies) or threshold inheritance, where the segregation can resemble simple Mendelian inheritance. If

a given hypothesis is to be accepted, test mating is in many cases simply waiting for more offspring to be born, in which the results are predictable. If the results deviate, a new hypothesis, which can be tested, has to be formulated.

How much data material is needed for a given test to give a final answer?

This is one of the important questions that have to be answered before the initiation of test mating.

Test mating of a bull, which is a suspected carrier, can be done by mating the bull with known affected cows, known carriers or the bull's own daughters. Taking these cases into consideration, how many normal offspring does the bull have to produce before he can be declared unaffected? The probability that one normal offspring is born in the three cases is respectively 1/2, 3/4 and 7/8. By applicating the multiplication rule for probabilities in mating with known carriers, the probability of getting one normal offspring is $(3/4)^n$. By means of this formula and a limit of significance of 0.05, 0.01 and 0.001 n can be determined to meet the classic probability criteria.

Number of normal offspring (n) which is needed to exclude a bull as carrier

```
Exclusion of bull                    Probability               Formula
as carrier by test          ----------------------------
mating with                 0,05    0,01      0.001       (x)ⁿ less than

Knowingly recessive           5       7        10         x=1/2
Known carrier                11      16        24         x=3/4
Own daughters                23
35        52          x=7/8
```

For normal offspring of a heterozygotic individual by father/daughters mating the probability is 7/8, as can be derived from Figure 5.6. The number of offspring needed to declare the father non-carrier on the 5% level is $(7/8)^n = 0,05$ corresponding to $n*\ln(7/8) = \ln(0,05)$, which rounded is n=23.

Mating with own daughters also has the advantage that the bull is tested as carrier of all possible recessive genes, not only the one he is suspected of having.

Figure 5.6 Theoretical segregation of recessive genes in test mating with own daughters.

Chi-square tests

By testing for segregation ratios for known mating a Chi-square test can be utilized. The test for mating with known carriers resembles the test for H-W equilibrium, but here the expected numbers are in the ratios 1:2:1 for co-dominant and 3:1 for dominant cases. The expected numbers are calculated by multiplying the expected frequencies (ratios) with N, as shown in the table below.

Segregation by test mating of Aa x Aa, codominant inheritance.

```
---------
Genotype          AA              Aa              aa        Total
Numbers, obs      30              51              39        = 120 = N
Frequency, exp    1/4             1/2             1/4       = 1,00
Numbers, exp      30              60              30        = 120
```

```
Deviation            0                -9              9
Chi-squared          0                1.35            2.70    = 4.05
----------
```

Df = 3-1 = 2, since the material only supplies the parameter N to the calculation of expected numbers. The Chi-square value is less than the H_0 test value of 5.99, which means that there is no statistical significant deviation from a 1:2:1 segregating on the 5% level.

Segregation by test mating of Aa x Aa, dominant inheritance

```
----------
Genotype        AA+ Aa          aa          Total
Numbers, obs      81             39          = 120 = N
Frequency, exp    3/4            1/4         = 1,00
Numbers, exp      90             30          = 120
Deviation         -9              9
Chi-squared       1.00           2.70        = 3.70
----------
```

Df = 2-1 = 1, since the material only supplies the parameter N to the calculation of expected numbers. The Chi-square value is just below the H_0 test value of 3.84, which means that there is no statistical significant deviation from a 3:1 segregation on the 5% level.

Applet for calculating the Chi-square for Mendelian segregation by test mating, is found here.

5.7 Field data, statistical tests

When the use of field data is needed, and it concerns newly occurred diseases, only families with affected offspring can be considered. When a segregation ratio for Mendelian inheritance is calculated on the base of such observations, it is necessary to correct for families that due to random chances are not segregating.

If recessive inheritance occurs and both parents in a family are heterozygotes the chance of the family not being detected, by means of an affected first born, is 3/4. After the birth of the second offspring the chances will be $(3/4)^2 =$ 9/16. Figure 5.7 shows how heterozygotic parents will segregate in a family of two. 9 out of 16 families will have no affected offspring, so by excluding these families in the final analysis, the classical Mendelian analysis becomes meaningless.

Figure 5.7
Theoretical segregation of heterozygotic parents in 7 out of 16 families of two.

In the following are formulas for testing and calculating the corrected segregation in families which have chances of segregation. Therefore the relevant data (shaded area) from Figure 5.7 is:

```
T = 14    -total number
A = 8     -number of affected
A₁ = 6    -number of families with 1 affected
A₂ = 1    -number of families with 2 affected
```

$T = 14$ -total number
$A = 8$ -number of affected
$A_1 = 6$ -number of families with 1 affected
$A_2 = 1$ -number of families with 2 affected

Figure 5.8 shows a path diagram of a mink family from a farm with an outbreak of tyrosinaemia. One of the clinical symptoms is, as shown in the picture, the fact that the eyelids stick together. The whelps die after one or two days with clinical symptoms. The cases occur when the whelp is about six weeks. A description of the disease can be found in Christensen et al. *Canadian J. Comparative Medicine*, 43:333-340, 1979.

Figure 5.8. Segregation of tyrosinaemia in standard mink (filled symbols), The photo on the right shows an affected whelp with the typical sticky eyelids.
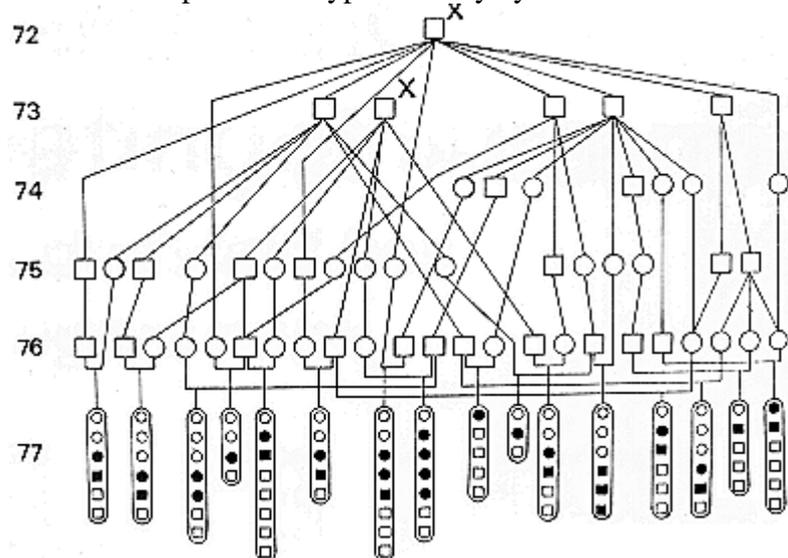
The formulas to apply according to "Singles method" for a test of Mendelian inheritance (for the proportion p-hat) are shown below. The actual numbers from the tyrosinaemia case, Figure 5.8, are inserted.

$$\hat{p} = \frac{A - A_1}{T - A_1} = \frac{32-4}{94-4} = 0{,}311$$

$$Var(\hat{p}) = \frac{T - A}{(T - A_1)^3}\left[A - A_1 + 2A_2 \frac{T - A}{T - A_1}\right]$$

$$= \frac{94-32}{(94-4)^3}\left[32-4 + 2*9\frac{94-32}{94-4}\right]$$

$$= 0{,}0034$$

$$Z^2 = \frac{(\hat{p} - p)^2}{Var(\hat{p})} = \frac{(0{,}311-0{,}25)^2}{0{,}0034} = 1{,}10$$

T and A are respectively the total number of offspring and the number of affected offspring. A_1 and A_2 are the number of families with respectively 1 and 2 affected offspring. Z^2 is Chi-squared distributed with 1 degree of freedom. P-hat is tested against the expected segregation frequency (p) which in the tyrosinaemia case is 0.25, as it is expected that both parents are carriers.

In the mink family is found that:

```
T = 94    -total number
A = 32    -number of affected
A₁ = 4    -number of families with 1 affected
A₂ = 9    -number of families with 2 affected
```

The calculation is carried out by inserting the actual numbers in the formulas above. There is no significant statistical deviation from a 3:1 segregation ratio as the test value is less than 3,84 (H_0: p(hat) = p =0,25). Families with more than 2 affected individuals are not directly part of the calculations, but contribute only to the total number of affected.
An applet for the calculation can be found here

If the number from the first theoretical example (Figure 5.7) is inserted in the formula for p(hat) we get (8-6)/(14-6) = 0.25, which is of course the correct result, as the data originated from mating with known heterozygotes.

Chapter 6. Definition of a quantitative trait, breeding value and heritability

6.1 Definition of a quantitative trait

In section 1.4 was given a definition of the terms phenotype and genotype for a quantitative trait, this equation is shown below. Until now the terms have only applied to qualitative traits.

$$\text{Phenotype} = \text{Genotype} + \text{Environment} \quad P = G + E$$

The phenotype can be seen or measured, in contrast to the genotype which can be transferred to offspring. In chapter 1 the distribution of fat-% in the milk in two Danish dairy breeds was given as an example of a quantitative trait.

The distribution of fat-% in the milk is characterized by a mean value and a standard deviation. For HF dairy breed the mean value is 4.3 % and the standard deviation is 0.25 % units. Most cows have a fat-% of around 4.3, and only very few have a fat-% of 3.5 or 5.0. For the Jersey breeds the mean fat-% is 6.4.
The phenotype of an animal should always be evaluated as being a deviation from the mean value of the population. For instance a Jersey cow should be evaluated based on the mean value of the Jersey and not on the mean value of HF.

Quantitative traits are normally influenced by the effects of several gene pairs and the environment, and have a normal distribution of population values. This is **polygenic** inheritance, as opposed to single gene (monogenic) inheritance or Mendelian genetics, which have been described in the preceding chapters.

6.2 The terms genotype value, breeding value and dominance deviation

The phenotype value (P) of an animal can be measured and evaluated as a deviation from the mean value of the population, P-bar.

The genotype value (G) of an animal equals the phenotype mean value of individuals with the same genotype.

The definition of the breeding value (A) of an individual is based on an infinite number of offspring as two times the deviation of the offspring's mean value from the mean value of the population, when random mating is used. I.e. all offspring are half sibs, see Figure 6.1.

Figure 6.1
The definition of breeding value (A) of a male is based on an infinite number of offspring, all being half sibs.

$$A = 2\,(\bar{P}_g - \bar{P})$$
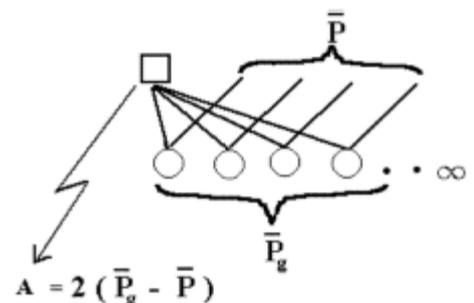
The mean value of the population can be added if the breeding values should be on a normal scale. The definition is valid when it concerns both a specific trait, and the effect of genes on a single locus. The formulas corresponding to the definitions applied on data from a single locus are shown in Figure 6.2. It is a precondition that the locus has an effect on the examined trait and that the average of the animals (with different genotypes) differs from each other. The average of the population can be calculated as a mean value. I.e. the sum of the three genotypes multiplied by their respective frequencies. See the example of weight of mice shown in Figure 6.2.

Figure 6.2.
Definition of breeding value for the genotype A_1A_1 and it is exemplified by weight of mice depending on the genotypes in a locus. A male with the genotype A_1A_1 gets an infinite number of offspring after random mating in a population.
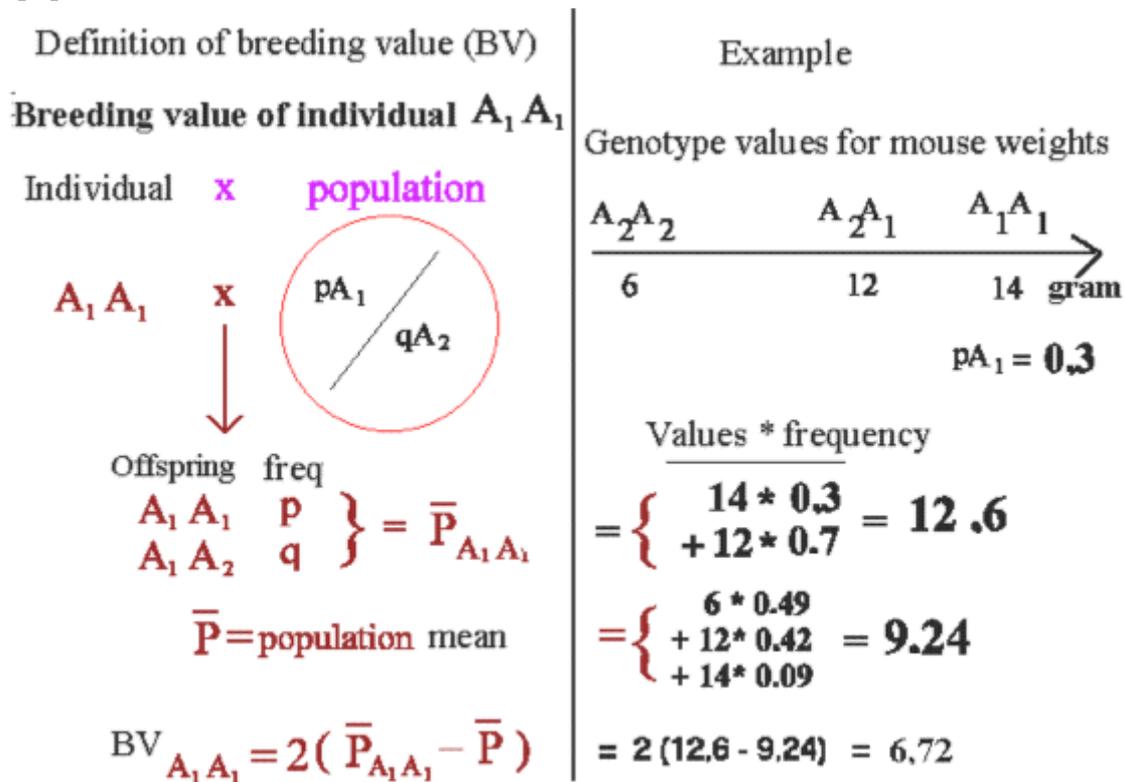
When an animal A_1A_1 is subjected to random mating in a population, the animal can get two types of offspring, A_1A_1 and A_1A_2, as seen in Figure 6.2. The gene from the population determined the genotype in the offspring. The frequencies of the two possible genotypes are equal to the population frequencies p and q.

The mean value of a population or of A_1A_1's offspring is calculated as:

$$\text{Weighted mean or}$$
$$\text{Mean value} = \Sigma \text{ Value} * \text{ frequency}$$

In the example of the weight of mice (Figure 6.2), calculation of the mean value of the population (P bar) and breeding value for the genotype A_1A_1 is shown. A similar breeding value can be calculated for genotype A_2A_2, which, by random mating in the population, gets offspring of the types A_1A_2 and A_2A_2 with the frequencies p and q. These results are derived from Figure 6.2 by mating the population with A_2A_2 instead of A_1A_1. Thus the breeding value of A_2A_2 is 2(12*0.3+6*0.7 - 9,24) = -4,88.

The breeding values of the heterozygotes is the average of the two pure genotypes, as half of its gametes is type A_1 and the other half is type A_2. The breeding values are often called the additive values, because they are proportional to the number of A_1 genes in the genotype.

If the genotype value (G) has a more or less distinct dominance, the average value of a specific genotype is not equal to the breeding value (A). But there will be a rest which is caused by dominance deviations (D). The degree of dominance is determined by the value of the heterozygote type compared to the two homozygote types:

- No dominance - the heterozygote is equal to the average of the two homozygotes
- Complete dominance - the heterozygote has the same genotype value as one of the homozygotes
- Overdominance - the heterozygote is out side the homozygotes

An example of calculation of breeding value (A) and dominance deviations (D)
Below is the results of an example of variation in transferrin locus. In Jersey cows the genotype tt yield 2082 kg. milk and cows with the genotypes Tt and TT yields 1882 kg. milk. Below are the actual numbers, the calculation of the mean value is based on the gene frequencies p and q, which are respectively 0.67 and 0.33.

```
Genotype                       TT        Tt              tt
-------------------------------------------------------------
Kg Milk              1882              1882            2082
(Genotype)frequence        p²=0.45        2pq=0.44      q²=0.11

And mean value   = 0.45*1882 + 0.44*1882 + 0.11*2082 = 1904 kg
```

The breeding values are calculated as deviations from the mean values of the population, see scale:

```
          TT and  Tt    Mean value                tt

          1882    1904                             2082    Original scale
          ---|-------|--------------------------|-->  Genotype scale, kg
            -22      0                             178    Deviation from mean
value
```

```
Genotype    breeding value
---------------------------------------------------
             TT   p       Tt    q
TT        2*[(-22*0.67 + -22*0.33) – 0] = -44
Tt                                       = 22,6 ,mean value of the homozygotes
tt        2*[(-22*0.67 + 178*0.33) – 0] = 89.2
---------------------------------------------------
```

By random mating in the population the genotype TT gets offspring which are of types TT and Tt with the frequencies p and q. This can be derived from Figure 6.2.

The genotype values are determined as deviations from the population mean (see scale). And the breeding values are calculated according to the definition. Finally the dominance deviations (D) can be determined as a rest as shown below.

```
Genotype     G   =    A    +   D
------------------------------------
TT          -22  =  -44    +   22
Tt          -22  =   22,6  +  -44.6
tt          178  =   89,2  +   88.8
------------------------------
-----
```

Results derived from using the applet on the example of mice weigths shown in Figure 6.2

It is possible to calculate the variance for a locus. The variance can be caused by the differences in breeding values or in dominance deviations. The variance (V) is calculated as a mean value of the squared breeding values. For instance, the variance of the breeding values: V_A = $(-44-0)^2*0.45 + (22.6-0)^2*0.44 + (89.2-0)^2*0.11 = 1926$, which, in this case, are the breeding values squared and multiplied by the genotype frequency. The average of the breeding

Genotypes	aa	Aa	AA
Meassured values	6	12	14
Mean value = 9.24	q	Calculate !	p
	0.700000	q + deltaq	0.300000
Genotypes	Genotypic values, G	Aditive values, A	Dominance deviations, D
AA	4.7600	6.7200	-1.9600
Aa	2.7600	1.9200	0.8399
aa	-3.2399	-2.8799	-0.3600
	VarT = 10.3823	VarA = 9.6767	VarD = 0.7056

values can be calculated as a mean value $-44*0.45 + 22.6*0.44 + 89.2*0.11 = 0$, which was to be expected.

An applet for calculating G, A and D are shown in the figure above and found here
Experiments to determine quantitative gene effects are described in chapter 12. Loci, which has effect on a quantitative trait, are often called QTL's (quantitative trait loci).

6.3 The terms additive variance and heritability

The term **additive variance** refers to the variance of breeding values (V_A). (In some of the next chapters the symbol sigma^2 is used synonymous with the variance.) Section 6.2 showed how the breeding value for each genotype in a locus could be estimated. When the genotype values are known, the additive variance for the locus can also be calculated. If independence occur between the loci, which have additive effect on a trait, then the joint additive variance will be the sum of variances from each locus. For individuals in a population the joint additive variance cannot be calculated directly, since the breeding value of the individuals is not known. It is not possible for all individuals to have the infinite number of offspring, which are required when the definition of breeding value is used. But V_A can be calculated indirectly, as shown later in this chapter.

The **phenotypic variance** (V_P), on the other hand, can always be calculated using the phenotypic values of the individuals in the population.

The **heritability** (h^2) or the degree of heritability for a trait is defined as V_A/V_P. The heritability can vary between 0 and 1, since the additive variance is only part of the phenotypic variance.

Even though the additive variance cannot be directly estimated, the heritability can be estimated on the basis of the similarity between related individuals, which is measured either by a correlation or a regression. For instance, the regression of offspring on one of the parents. The heritability is estimated as the ratio of the calculated correlation coefficient and the coefficient of relationship. Which means that the maximum expected correlation between parent and offspring is 0.5. If the correlation coefficient is higher, it means that the similarity is caused by other factors besides the genetic.

The best method for estimating the heritability is the use of the correlation between half sibs, since they normally have only genes in common. Full sibs, on the other hand, often have a maternal environment in common too, which causes the litter mates to be even more alike, and have more in common than genes. In such cases the heritability would be overestimated.
The parent-offspring correlation is seldom used in practise, since this method needs data from two generations, often with different environment. It can also be a problem if both parent and offspring live in the same environment.

Estimation of the variance component, caused by common environment for full sibs, can be carried out by comparing the estimated correlation based on respectively half and full sibs. The correlation between full

sibs should be twice the one for half sibs. Often it is higher, the reason for this is that the full sibs have a common environment. This component is assigned the symbol c^2, c for 'common environment'. The relative phenotypic variance has a genetic part (h^2), a common environmental part (c^2), and a third part. This remaining part of the variance is caused by random environmental effects, and gets the symbol e^2. The random environmental part is for most traits the largest part of the phenotypic variance. Below is shown a table of the defined parameters.

```
Additive variance            V_A
Phenotypic variance          V_P
Heritability                  h²
Common environment            c²
Random environment            e²
Half sib correlation         (1/4)h²
Full sib correlation         (1/2)h² + c²
```

Historically, the heritability has another interpretation. By selecting animals (shades area in Figure 6.3), which deviate from the mean value, the average of the next generation will change. The change is proportional to the difference in selection and the heritability following the formula:

$$R = h^2 * S$$

The selection response (R) equals the heritability multiplied by the selection difference (S). The heritability is 0, if the trait remains constant by selection, and it is 1, if the changes are equal to the selection difference, see Figure 6.3.

6.4 Estimating the heritability and common environmental effect

Selection experiment for estimation of heritability.

In Figure 6.5 is shown a selection experiment for September weight in mink, carried out by G. Lagerquist, the Swedish Agric. Univ. published in J.Amim.Sci., 1993,71:3261-72. Only the males were selected with S equal to around 225 gram in each generation - the upper curve in the figure. The lower curve in the figure shows the changes for each generation, R, that is equal to around 50 gram per generation. The heritability is estimated by
$R = h^2 S/2$, as only the males were selected:
and inserted, $50 = h^2 * 225/2$ which correspond to h^2 equal to 0.44.

$R = h^2 * S$ can also be interpreted as the regression of offspring on the average of the two parents, where the regression coefficient b correspond to h^2.

Figure 6.3.
Selection response (R) equals the heritability multiplied by the difference in selection (S). The change is seen in the offspring
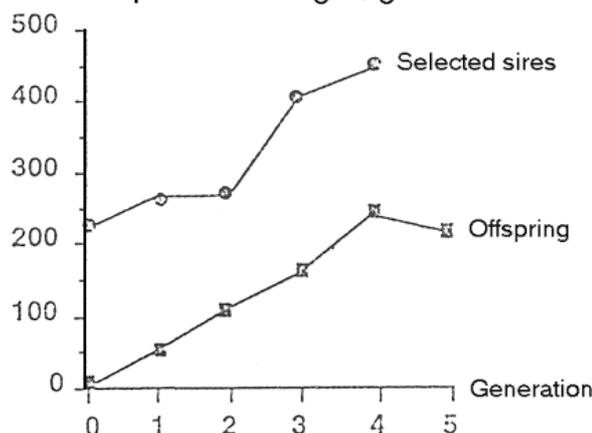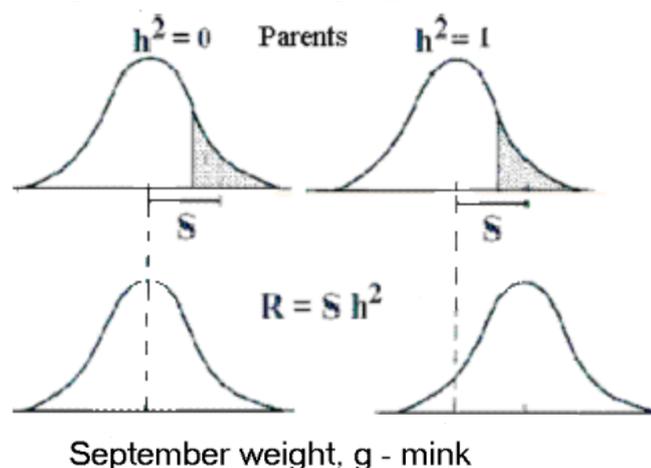
Figure 6.5. Example of selection experiment on which calculation of heritability can be based.

Heritability estimation based on calculated correlation.
The use of a selection experiment for estimating the heritability takes long time, at least several generations. Now will be presented methods for estimation of the heritability based on data from one generation only.

The heritability (h^2) can be estimated as the calculated correlation (r or t) between related individuals divided by the coefficient of relationship (a). This is correct if inheritance (genes) is the only cause of similarity between the involved individuals. The formulas are as follows:

$h^2 = r/a$, or $r = a*h^2$.

The calculation of the correlation can be done according to the classical method (see Figure 6.4). When considering several individuals in a group, as for instance half sibs, calculations are done by means of analysis of variance and estimation of the interclass correlation.

Estimation of the heritability can be based on full and half sib material as can be found in all multipare species. The heritabiliy measures what degree of similarity exists between related animals sharing identical genes.

If environmental factors are causing the similarity, this can be estimated as shown in the formula below. The common environmental factors are particularly important in the litter's first months, when the mother and its maternal abilities cause the special environment. In half sibs, having a common father, the common genes are usually the only cause of similarity.

The general formula for interpretation of correlation between related individuals is shown below, and as symbol of this correlation t is used, this is also the symbol of the interclass correlation.

Figure 6.4.
An example of a two-dimensional distribution for calculation of correlation, typically a trait of a mother and one of its offspring.

$t = a*h^2 + c^2$

Now the heritability can be calculated on the basis of half sib correlation, c^2 being 0. Then c^2 for full sibs can be calculated based on the full sib correlation.

Example: Weight of mink aged 8 weeks, sample size: 508 animals in 107 litters with 37 sires.
In a statistical analysis of the weight of 508 mink puppies was recorded.
Half sib correlation t = 0.03 and
Full sib correlation t = 0.41.

These numbers put into the genetic model results in:
Half sib correlation $t = (1/4)*h^2 + 0 = 0.03$ corresponding to $h^2 = 0.12$.
Full sib correlation $t = (1/2)*h^2 + c^2 = (1/2)*0.12 + c^2 = 0.41$ corresponding to $c^2 = 0.35$. For calculation of c^2 for full sibs the heritability calculated for half sibs is applied.

There is a very low heritability of growth in this early age, where the maternal factors play a large

The material for calculation of heritability of height of humans was shown in chapter 1. As mentioned there, the heritability was calculated as the regression of offspring on basis of the average of the parents. For height of humans it was around 0.6.
Mother-offspring correlation was found to be 0.4 while the correlation between father and offspring was only 0.2. This lower correlation between father and offspring might be caused by the fact that these do not share the same significant common environment as the mother and offspring. The 0.2 correlation between

father and offspring gives a heritability of 0.4 for height of humans. This means that the mother-offspring correlation is highly influenced by their common environment.

For data and the applied SAS program for height of humans, material from 2000, click here.

Further examples of application of the analysis of variance for calculating the repeatability coefficient for litter size of mink and heritability for the body length of swine. For the SAS program and the mink material, click here and for the swine material, click here

Chapter 7. Estimation of breeding values

7.1 Estimation of breeding values, general

In a large randomly mated population, each individual should on average give birth to two offspring in order to maintain the size of the population. The distribution of the number of offspring in the population has a left skewed binominal distribution (Poisson distributed) with a average value of 2 and variance of 2. Which means that the number of offspring per individual can vary from 0 and upwards, the values 0,1,2,3,4 and 5 being the most frequent. The exact breeding value, based on the definition given in chapter 6, cannot be calculated in such a population because of the small number of offspring. An estimation is all we can get.

An estimated breeding value is often called an index (I). The index can be estimated on the basis of information of phenotype values from all possible relatives. A simple regression line or multiple regression can be used. The higher the number of relatives is the better the estimation will be. Correlation between the true breeding value (A) and the index is given the name **Accuracy** and it has the symbol r_{AI}.

The estimated breeding value is based on a theory of linear regression and correlation. The basic definition of these terms can be seen here, or more detailed in statistical textbooks.

7.2 Formulas for calculating estimated breeding value based on uniformly related phenotypes

Figure 7.1 shows the relationship between phenotype measurements of a uniformly related group of individuals, P's, and the breeding value index (I), which is an estimate of the true breeding value (A). Additionally the figure shows the significance of the accuracy (r_{AI}), as the accuracy squared for the index is equal to a reduction of the variance of the estimated breeding value. All values can be interpreted by use of the classical regression equation:

P's = phenotype values of the trait
n = number of measurements
P_g bar = average of a uniformly related group of P's
a' = degree of relationship between the P's and the animal being estimated for the index
a = degree of relationship between P's
P bar = average of the population
A bar = average breeding value of the population = P bar
h^2 = heritability
c^2 = common environmental factor for the P's
$t = a*h^2 + c^2$ as defined in section 6.4

Figure 7.1.
Estimation of breeding value indexes based on regression of phenotype measurements (average of a uniformly related group of P's). The definition of the parameters, see above. An example of use of the formulas is given in Figure 7.2.

$$I = \bar{A} + \frac{n \cdot h^2 \cdot a'}{1+(n-1)\,t}(\bar{P}_g - \bar{P})$$

$$\bar{P} = \bar{A}$$

$$\frac{\Sigma P}{n} = \bar{P}_g$$

$$(\bar{P}, \bar{A})$$

$$\sigma^2_{A.I} = \sigma^2_A(1 - r^2_{AI})$$

$$r^2_{AI} = \frac{n \cdot h^2 \cdot a'}{1+(n-1)\,t}\; a'$$
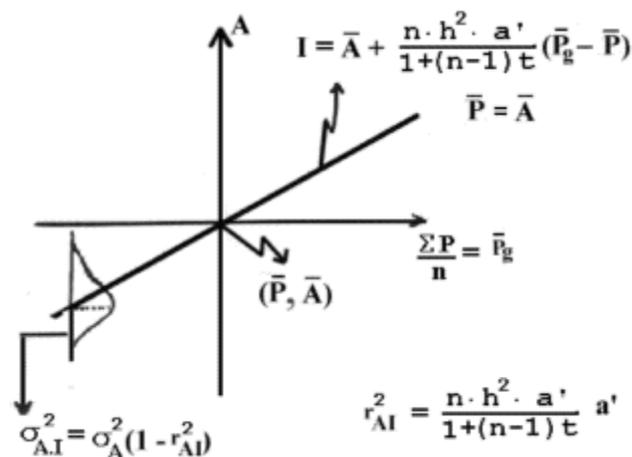
The estimated breeding value or the index (I) can be calculated by means of the formulas given in Figure 7.1. They can be used in all possible combinations of relationship. It only requires uniform relationships. The index depends on the two coefficients of relationship (a' and a), the number of measurements (n), the heritability (and common environment) and the phenotypical average (P_g bar) of the phenotypes. The accuracy (r_{AI}) of the index does not depend on the last factor.

Figure 7.2 gives an example. It is a genealogical diagram with the relevant coefficients of relationship for calculation of an estimated breeding value for milk yields for a bull based on the records of 20 progenies being half sibs (a progeny test). The genealogical diagram is made so that the phenotype measurements, the P's, have an blank circle. For the bull being examined the square is filled in.

Figure 7.2
shows a diagram of relationship and a calculation of estimated breeding value for milk yield for a bull based on 20 half sib daughters with an average milk yield of 7500 kg.

Breeding value index for ■ based on n measurements ○

Example: 20 daughters with an average of 7500 kg milk

a'=0,5   a and a' coefficients of relationship

a=0,25

$\bar{P}_g$ – Average measurements
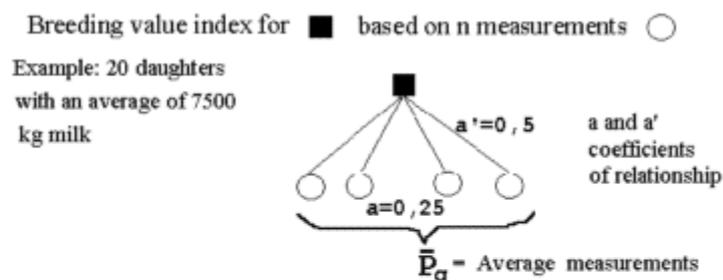
$$h^2 = 0,30 \ , \quad c^2 = 0 \ \text{og} \quad \bar{P} = 7000$$

$$I = \bar{A} + \frac{n \cdot h^2 \cdot a'}{1+(n-1)\,t}\,(\bar{P}_g - \bar{P}) = 7000 + \frac{20 \cdot 30 \cdot 5}{1+19 \cdot 30 \cdot 25}\,500 = 7618$$

$$r^2_{AI} = \frac{n \cdot h^2 \cdot a'}{1+(n-1)\,t}\,a' = \frac{20 \cdot 30 \cdot 5}{1+19 \cdot 30 \cdot 25}\,0,5 = 0,618$$

Below are three additional examples of diagrams of relationship defining the coefficients of relationship of uniform relationship. Additional diagrams would be superfluous since the parameters are here given for additional examples.

Figure 7.3.
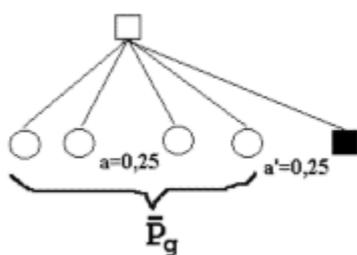
diagram of index based on half sibs.
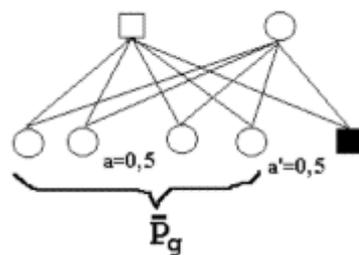
diagram of index based on full sibs. Both parents have the same relationships as a full brother. Therefore, the index is the same for a brother or either of the two parents.
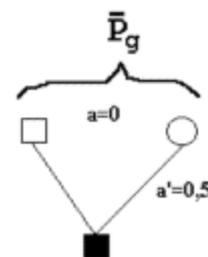
diagram of index based on the parents.

In the table below are given the parameter values of several examples of index calculation in cases concerning uniform relationship conditions. The relevant coefficients of relationships are put into the formula. And when an infinite number of observations are present, the limit value is placed as the regression coefficient and the squared accuracy.

Table Breeding value estimates from uniform measurements of n observations with mean value $\overline{P}_g$ using the formulas. Infinite is abbreviated inf.

A	$\overline{P}_g$	n	a'	a	t	$I = \overline{A} + \dfrac{a'nh^2}{1+(n-1)(ah^2+c^2)}(\overline{P}_g-\overline{P})$	$r^2_{AI} = \dfrac{a'nh^2\,a'}{1+(n-1)(ah^2+c^2)}$
Own	none	0	-	-	-	" + 0	0
"	own	1	1	-	-	" + h^2	h^2
"	"	2	1	1	h^2+c^2	" + $2h^2/(1+h^2+c^2)$	$2h^2/(1+h^2+c^2)$
Offspring	both parents	2	.5	0	-	" + h^2	$.5h^2$
Offspring	grand parents	4	.25	0	-	" + h^2	$.25h^2$
Father	offspring	1	.5	0	-	" + $.5h^2$	$.25h^2$
"	"	inf.	.5	.25	$.25h^2+0$	" + 2	1
Half brother sib	half	inf.	.25	.25	$.25h^2+0$	" + 1	.25
Full brother sib	full	inf.	.5	.5	$.5h^2+c^2$	" + <1	$<.5$

The table parameters are defined in relation to Figure 7.1. Any animal, having the same coefficient of relationship, a', as the measured group, would also have the same estimated breeding value. A full brother, the father or the mother, for instance, have a' = .5 to the same full sib. Or a father, evaluated through an offspring, has the same formula as a son evaluated through the father.

Line 1 in the table corresponds to the formula having 0 measurements, and the index equals the population mean with an accuracy of 0. This presents the 'basis' for evaluation of all the other indexes.
Line 2 in the table correspond to a very common situation. This is normally called a phenotype test. The animal's breeding value is based on measurements of the animal - the equation is:
$I = A(bar) + h^2(P - P(bar))$ with a squared accuracy of h^2.
Line 3: Two measurements of the animal, which could be litter size in first and second litter. The correlation between repeated measurements is called **the coefficient of repeatability**. This always contains some common environmental effects for the two measurements.
Line 4 shows the relationship between the average of the parents and the offspring. In section 6.4 this relation was used to estimate the heritability. Therefore it is no surprise to see the same formula again.
Line 7 shows the evaluation of a father based on an infinitely large number of offspring. This corresponds to the definition of breeding value and the weight factor equals 2 and the accuracy squared is equal to 1, this line correspond to the definition of breeding value given in chapter 6.

Accuracy of estimated breeding value

Figure 7.4 shows the accuracy of progeny testing based on n offspring. From the curves it becomes evident that there is larger accuracy at higher heritabilities than at low ones. An increase in the number of offspring can compensate for the low heritability.

Some general remarks should be added about the accuracy of estimated breeding value. By scrutinizing the formula for accuracy it becomes clear that:

1) For traits with high heritabilities the phenotype of the individual is a good source of information.

2) For traits with low heritabilities the use of several measurements can compensate for poor accuracy in a single measurement. This is only possible when large groups of offspring of full and half sibs are measured.

Figure 7.4. Accuracy of progeny testing - depending on the number of offspring and the magnitude of the heritability

$$\sqrt{\frac{n\,h^2 a'^2}{1+(n-1)\,t}}$$

$$a' = 0.5$$

$$t = 0.25\,h^2$$

Below is a table and the appropriate action to take depending on the magnitude of the heritability

```
Heritability   Size              Example                    Selection
-----------------------------------------------------------------------
High           Greater than 45%  Fat deposition, milk fat %  Individual
Low            Less than   10%   Reproduction traits         Progeny test
Middle                     15-40% Growth                     Mixed
-----------------------------------------------------------------------
```

An applet for calculation of the most common cases of estimated breeding value is found here, and an example of the use of it is shown below. From the example it becomes clear that an offspring group of 20 half sibs makes a useful source of information.

BV-estimation		Mean value	Heritability %	Common Milieu %	Repeatability %
Calculate		100	25	5	30
Do Nullify BV etc.		Number	Average	Breeding Value (BV)	Accuracy**2 (r**2)
On offspring, full sib		20	110	105.7803	0.2890
On offspring, half sib		20	110	111.4285	0.5714
On own performance		5	110	105.6818	0.5681
On half sibs		100	110	108.6956	0.2173
On both parents		2	110	102.5000	0.1250
a' og a	.5 1.	2	110	101.9230	0.0961

7.3 Direct update of breeding values

As a basic rule for direct updating 3 types of information on an animal breeding value exist; the parents, own performance and the offspring. The pieces of information derived from half and full sibs as well as from grandparents are all included in the parent's breeding value. This is both a strength and a weakness in the

direct updating method as the parents can only be included once. Therefore, new information on half and full sibs cannot be utilized. Usually this is unimportant, so it is an advantage in several cases not to bother about this information.

The information, gathered from the 3 different sources for the use of estimating the breeding value, has the following forms. The heritability is h^2 and common environmental variance c^2. The estimated breeding value on an absolute scale has the symbol A hat. The formulas for combining the indexes (measured as deviations from the average) from the 3 possible sources can be seen below.

Parents:

$$\hat{A} = (\hat{A}_{Dam} + \hat{A}_{Sire})/2 \ , where \ I_1 = \hat{A} - \bar{A} \ \ and \ \ r^2_{AI_1} = (r^2_{AI,Dam} + r^2_{AI,Sire})/4$$

Own performance (n observations with mean value $\bar{P}_g$ and repeatability $r = h^2 + c^2$):

$$\hat{A} = \bar{A} + \frac{nh^2}{1+(n-1)(h^2+c^2)} (\bar{P}_g - \bar{P}) \ ,where \ I_2 = \hat{A} - \bar{A} \ \ and \ \ r^2_{AI_2} = \frac{nh^2}{1+(n-1)(h^2+c^2)}$$

Half or full sib offspring (n number of offspring with mean value $\bar{P}_g$):

$$\hat{A} = \bar{A} + \frac{.5nh^2}{1+(n-1)(ah^2+c^2)} (\bar{P}_g - \bar{P}) \ ,where \ I_3 = \hat{A} - \bar{A} \ \ and \ \ r^2_{AI_3} = \frac{.25nh^2}{1+(n-1)(ah^2+c^2)}$$

where a=.25 and .5 for half and full sib offspring, respectively.

The indices I_1, I_2 and I_3 can be combined in a new index I_c

$$I_c = b_x I_x + b_y I_y \ \ where \ x=1,2,3 \ or \ c \ and \ y=1,2,3 \ or \ c \ and$$

$$b_x = \frac{1-r^2_{AI_y}}{1-r^2_{I_x I_y}} \ , \ \ b_y = \frac{1-r^2_{AI_x}}{1-r^2_{I_x I_y}} \ and \ r^2_{AI_c} = \frac{r^2_{AI_x} + r^2_{AI_y} - 2r^2_{I_x I_y}}{1-r^2_{I_x I_y}}$$

There is only one condition, that is, the two indices must come from sources that only are related though the animal having its breeding value estimated. This also implies that

$$r^2_{I_x I_y} = r^2_{A.I_x} \cdot r^2_{AI_y}$$

An example of this condition: an index has been made of full sib offspring and is included in the combination index. A new index of additional full sibs cannot be included without violating the rule, as they are related to the sibs in the first index, through both the father and the mother. Half sib offspring can, on the other hand, be added when new information arrives, since they are only related through the animal in question.

For an applet containing examples and use of direct update of breeding values, [click here](#)

7.4 Estimation of breeding values in animal breeding, and use of gene markers

For large domestic animal breeding organizations estimation of breeding values is an enterprise which can give the members a large economic return. Therefore, very refined methods are applied to improve the index estimation, also see lections 12.1-2. In addition to what is shown here in terms of simple estimation of breeding value, it is of great importance to corrected for significant environmental factors, as for instance

seasons and herd average. It is also important that all related individuals contribute with information to the indexes.

Estimation of the breeding value is based on large equation systems for all animals in the population simultaneously. This provides the possibility of simultaneous correction of environmental effects.

Estimation of the breeding value of an animal can be calculated based on all possible combinations of information. Section 7.2 only shows examples in which uniform relationship exists for the phenotype measurements. Her only a few exceptions from this form shall be shown. The first one is estimation of the breeding value based on the estimates of the parents' breeding value, which is which is also shown in section 7.3:

```
I_offspring = (I_sire + I_dam)/2          and
r²_AI,offspring = (r²_AI,sire + r²_AI,dam)/4
```

Below is shown by formulas to calculate how much a **gene marker** can add to the information from a phenotype measurement. The trait's heritability is h^2 and the gene marker represents a part of the joint heritability, which corresponds to $(h_1)^2$.

When M is the marker effect and P the phenotype value, the estimated breeding value (I) is as follows:

$$I = \bar{P} + [(1-h^2)/(1-h_1^2)] \times [M - \bar{P}]$$
$$+ [(h^2-h_1^2)/(1-h_1^2)] \times [P - \bar{P}]$$
$$r^2_{AI} = [(h_1^2/h^2) + h^2 - 2h_1^2]/(1-h_1^2)$$

In the table below are results from the use of the formula for accuracy with and without a gene marker (large effect, 20% of the genetic variation, column +M).

	−M	+M	−M	+M	−M	+M	Transferrin locus
h^2	.05	.05	.25	.25	.50	.50	| .33
$(h_1)^2$	0	.01	0	.05	0	.10	| .01
r_{AI}^2	.05	.24	.25	.35	.50	.56	| .34

It appears that, for traits with low heritability, a significant improvement of the accuracy can be obtained by using the information on the gene marker. For traits with high heritability the improvement is very low.

Example: Index of a cow with a yield of 2300 kg milk and Tt in transferrin type from the population described in section 6.2. This cow becomes like the marker genotype (M=1882 kg milk) and its mean value is ($P_{ave.}$=1904).

The formula above is applied and h^2 and h_1^2 are taken from the table

```
I = P_ave + X*(M - P_bar ) + Y*( P - P_bar ),where

X = (1-h²)/(1-h₁²)   = (1 - 0,33)/(1 - 0,01) = 0,68

Y = (h²-h₁²)/(1-h₁²) = (0,33 - 0,01)/(1 - 0,01) = 0,32 by insertion is obtained

I = 1904 + 0,68*(1882-1904) + 0,32*(2300-1904) = 2016 kg milk
```

The accuracy as is seen from the table is only slightly improved by using the transferrin type in the index.

For estimating the effect of a gene influencing a quantitative trait (a QTL), reference can be made to lection 12.5.

Standardization of a breeding value index.
The breeding value indexes are normally expressed in absolute units. The index, for instance, of a bull with a daily weight gain of 1100 grams from birth to 12 months, many people have difficulties in deciding whether this is good or bad. The basis for the comparison is missing when the mean value is unknown. Therefore, by subtraction of the mean value the indexes are centered around zero. Thus the better half will have a positive index and the worse half will have a negative index.
Furthermore the indexes can be standardized both by subtracting the mean value and dividing it with its standard deviation. Such a population will have a mean value of zero and a standard deviation of one.
In Denmark standardized indexes are often used, so the mean value is 100 and a standard deviation of for instance 5. With such an index animals with values higher than 100 are better than the average and animals with an index value lower than 100 are worse than the average.

Chapter 8. Genetic changes by selection

8.1 Difference of and intensity of selection

Figure 8.1 shows a normal distribution of a quantitative trait. A limit of the selection can be defined according to the proportion being used for breeding.
S designates the difference between the population average and the mean value of the selected animals. If a specific group has been selected, its actual mean value is calculated and then S can be calculated. When a certain proportion of animals is used in breeding, the standardized difference in the selection can be seen in a table when using truncation selection. This difference is called **selection intensity**, as shown in Figure 8.1, it can be estimated as S/sigma$_P$, sigma$_P$ being the phenotypic standard deviation.

Figure 8.1 also shows the connection between estimated breeding value and the simple formula for selection response $R = h^2 * S$, given in section 6.4, and its relation to the estimation of breeding value by means of its own phenotype, cf. section 7.2.

Figure 8.1. Shows a graphic interpretation of the selection difference (S), the selection intensity (i) and selection response (delta G) by use of individual selection.

proportion of selected animals

$$\leftarrow 2\,\sigma_P \rightarrow \leftarrow S \rightarrow$$

$$S = \text{Selection difference}$$

$$i = \frac{S}{\sigma_P} = \begin{matrix}\text{selection}\\\text{intensity}\end{matrix} \left(\begin{matrix}\text{Use}\\\text{table}\end{matrix}\right)$$

$$(I - \bar{A}) = (h^2(\bar{p}_S - \bar{p}))$$

$$\Delta G = h^2 S = h^2 i\,\sigma_P = h\,i\,\sigma_A$$
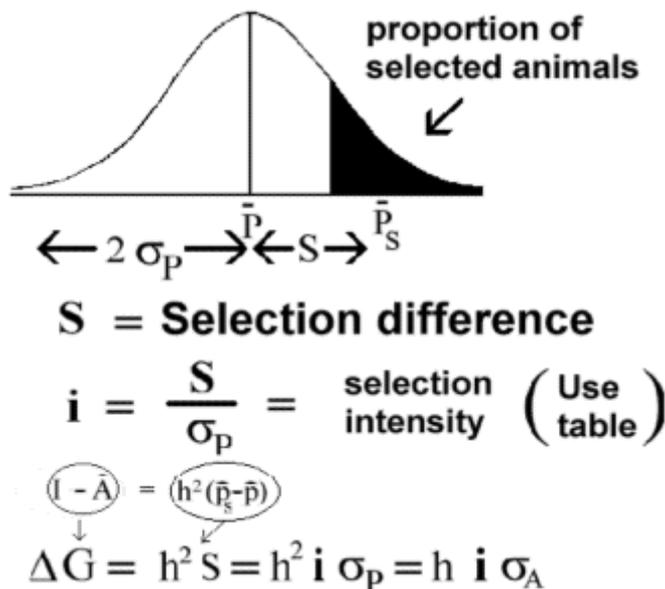
Below is a table of 'i' calculated on the basis of the proportion of animals selected for breeding.

Proportion of animals for breeding	Intensity i	Proportion of animals for breeding	Intensity i	Proportion of animals for breeding	Intensity i
1.00	0				
0.90	0,20	0.09	1,80	0.008	2,74
0.80	0,35	0.08	1,85	0.006	2,83
0.70	0,50	0.07	1,91	0.004	2,96
0.60	0,64	0.06	1,98	0.002	3,17
0.50	0,80	0.05	2,06	0.001	3,38
0.40	0,97	0.04	2,15	0.0008	3,43
0.30	1,14	0.03	2,27	0.0006	3,51
0.20	1,40	0.02	2,42	0.0004	3,61
0.10	1,76	0.01	2,67	0.0002	3,79

An example of the use of the table: If 30% of the animal is used for breeding, i is 1.14.

8.2 Selection response

Below is shown a general formula for expected genetic changes by selection. The changes are proportional to the intensity of the selection, the accuracy of the estimated breeding value and the additive standard deviation, sigma$_A$. The change take place from one generation to the next, which means that the average breeding value of the parents corresponds to the mean value of the population in the next generation. If the changes are to be measured in time units, they are inversely proportional to the generation interval.
The generation interval is defined as the average age of the animals when they become parents. In the human population the generation interval is said to be 30 years. I.e. the parent is on average 30 years old when they give birth to a child.

The general formula for selection response is shown below.

$$\Delta G_{year} = \frac{i \; r_{AI} \; \sigma_A}{L}$$

$i = $ Selection intensity

$r_{AI} = $ Accuracy

$\sigma_A = $ Genetic standard deviation

$L \; = $ Generation interval

The general formula for delta G fits well with the results for accuracy of individual selection, h, given in Figure 8.1. The general formula for selection response can be extended slightly when taking into account the different roles the two sexes play in forming the new generation. The formula is divided into two components, one for the sires' contribution and the other for the dames' contributions, see below.

$$\Delta G_{year} = \frac{i_\female r_{AI\female} \sigma_{A\female} + i_\male r_{AI\male} \sigma_{A\male}}{L\female \quad + \quad L\male}$$

An applet for calculating the selection response is found here

The formula for delta G is important for predicting what will happen if a specific breeding program is decided upon. Thus the model studies can be performed with an eye on design of an optimal balance between the generation interval and the intensity and/or accuracy of the selection. At short generation intervals fewer individuals are born to select among. This influences both the intensity and the accuracy negatively compared to a situation where longer generation intervals occur.

8.3 Selection of threshold traits

From chapter 1 shall be repeated that a threshold trait is inherited as a quantitative trait, containing the feature that it occurs as an either or trait. For instance are mastitis in dairy cattle or a heart anomaly in a new-born puppy threshold traits.

Whereas the number of outbreaks of mastitis in a cow can be regarded as a semi quantitative trait, particularly seen on the back-ground of an entire life-span with many lactations. Concerning the number of mastitis cases, there is no doubt, that some linearity occur between the number of cases and the animal's resistance to the disease.

In other cases it is more difficult to imagine a linear scale. For instance in the so called 'stick and feather test' for mink, designed to test their reaction patterns. The possible outcome of the test is aggressive, curious or fearful, all of which can be seen from the same test. Is it relevant to put the three classes on the same linear scale and if so how?

To get a meaningful result from such a test three independent either-or scales has to be used. 1) Aggressive or not aggressive, 2) fearful or not fearful, and 3) curious or not curious, see Figure 8.2.

Figure 8.2. The three characteristics aggressive, curious and fearful cannot be interpreted from a linear scale. Each of them has to have its own either-or scale.
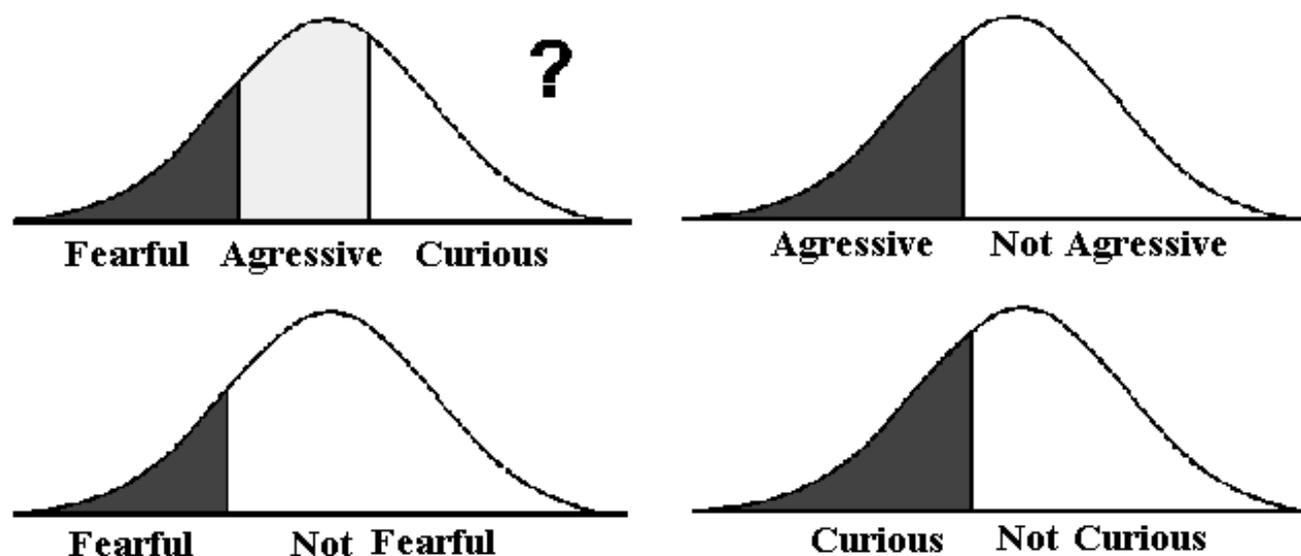
An example of the use of a semi-linear scale: Hip dysplasia (HD) in German Shepherd dogs from Andresen et al. HD index: Bedømmelsen bør bruges efter hensigten. Hunden dec.1994.
Figure 8.3 shows an example, which deals with Hip dysplasia (HD) in German Shepherds. Here is data from a population of 1406 X-rayed dogs. All the dog's hips were evaluated on a scale from E2, E1 to A2, A1, 10 classes all in all as shown in the figure. E2 dogs have a very high degree of HD, and A1 dogs have perfect hips. Dogs from B2 and below have HD in an increasing severity.

As can be seen in Figure 8.2, the scale is not linear when the data is fitted into a normal distribution. In the calculation of the HD index, class mean has been used instead of a completely linear ten step scale. The class mean has been determined from the normal distribution with a mean value of 0 and a standard deviation of 1.

Figure 8.3
A partly linear scale is used to graduate the severity of HD. The number in each class is adapted

 to the normal distribution.

Classes	E2	E1	D2	D1	C2	C1	B2	B1	A2	A1
Number	7	43	48	68	80	111	241	447	326	35
Middle	-2.8	-2.05	-1.65	-1.35	-1.05	-0.80	-0.45	0.20	1.10	2.25
HD-Index	38	55	64	70	77	82	90	100 104	124	150
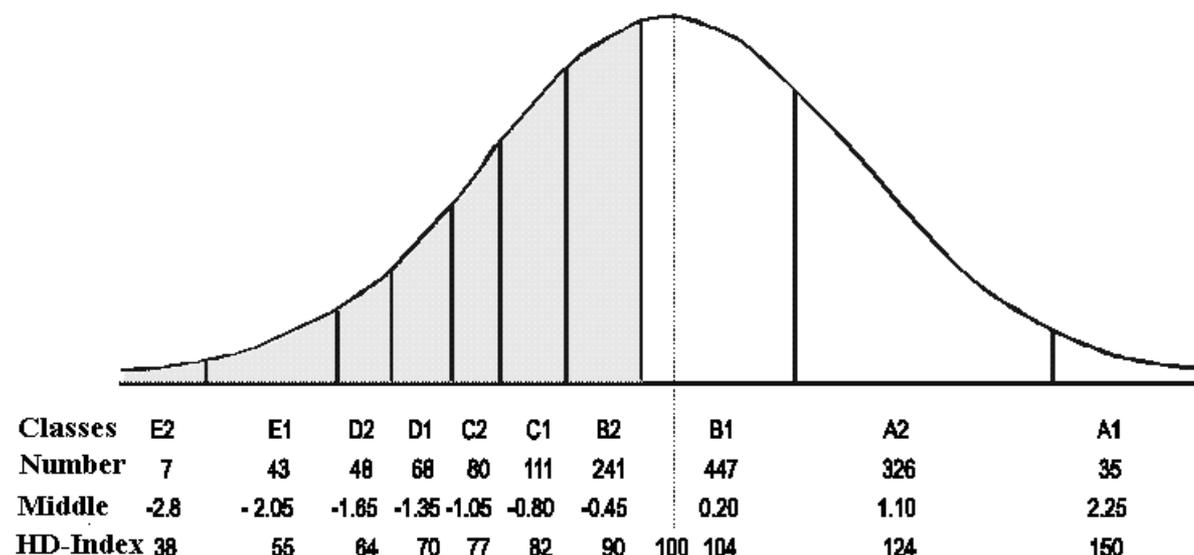
The estimation of the breeding value (HD index) of a male on the basis of 15 offspring is shown below. The offspring had the scores 4 A2, 9 B1, 1 B2 and 1 D1 with a mean value of .2933. To rescale the breeding values, 100 is added and the deviation is multiplied by 100. Animals with an index higher than 100 are better than the average. The actual calculation of the breeding value is done accordingly to the formulas given in section 7.2

$$\textbf{BV-estimation for HD}$$
$$\textbf{based on 15 progenies} \quad h^2 = 0,22$$

$$\bar{P}_g = \frac{4\cdot 1,10 + 9\cdot 0,20 + 1\cdot(-0,45) + 1\cdot(-1,35)}{15} = 0,2933$$

$$I = 100 + \frac{15\ h^2\ 0,50\ 100}{1+(\ 14\ )0,25\,h^2}\ (0,2933 - 0) = 127$$

The calculated HD-index lies significantly above the average, which can also be deducted from the fact that the male has only two offspring with B2 or below.

The calculated HD-index can be standardized to vary between 50 and 150. The present index is randomly 'standardized' with the factor 100.

Figure 8.4 shows X-rays of the best hips, A1, and the worst, E2. The pictures have been taken at Røntgen-Klinikken at KVL.

Figure 8.4

H-D score, left A1 and right E2.

From April 2000 The Danish Kennel Club has altered the scale for HD evaluation, it is now a five-step scale containing only the grades A, B, C, D and E, this new scale has been adapted internationally. In the present example the HD-index can be recalculated with the new class means 1.02, -.02, -.90, -1.47 and -2.16.

Comparison of occurrence of threshold and Mendelian diseases. In chapter 5 was given a detailed description of the segregation of Mendelian diseases within families. All these forms run within families. The same is true for a threshold disease.
For diseases with low population frequency it is not possible to discern between a Mendelian and a threshold disease. In both cases the frequency of the disease is much higher in individuals closely related to a diseased one, than the frequency in the general population.
The differentiation between the two forms of inheritance can only be done by means of test mating. In case of the Mendelian inherited diseases the exact segregation ratios can be predicted. This is not the case for threshold traits.

Estimating heritability for threshold traits. In chapter 6 methods for estimating heritabilities for normal distributed quantitative traits was given. The heritability can be estimated from a threshold disease when the population frequency is known, as well as the frequency in offspring of affected animals or relatives of affected animals.

Figure 8.5 shows an example of heritability estimation in graphic form.

Figure 8.5
Calculation of heritability based on disease frequency in relatives of affected animals. In this case affected parents.

The frequency in the population is 5 % and the frequency in offspring of affected animals is 20 %. The situation can be regarded as a selection experiment, where only affected animals are selected for breeding. Now the simple formula for delta G or the selection response, $R = S*h^2$ can be used. The threshold trait is measured in a standardized normal distribution with $S = i$. The response R is the difference between X_1 and X_2. (They can be looked up in a table of the normal distribution.) The selection intensity (i) is also in the table of i. The heritability in this case is estimated to 0,39.

If the frequency of affected animals is known, as for instance in first degree relatives (father, mother or full sibs), the calculations are carried out as shown in the example, but then $h^2/2$ is estimated as the R/i. In data from second degree relatives, as for instance half sibs, the estimated result is $h^2/4$.

Basic formula
$$R = h^2 S = h^2 i$$

$$h^2 = \frac{R}{i}$$
$$= \frac{X_1 - X_2}{i}$$
$$= \frac{1{,}6 - 0{,}8}{2.06} = 0.39$$

A statistic of the present type can also be regarded as an epidemiological investigation, the risk factor is relationship to an affected individual. The relative risk of getting the disease is 4 times larger in offspring of two affected parents than in a randomly chosen individual in the population shown in Figure 8.5.

[An applet for calculating heritability of threshold traits](#)

8.4 Genetic correlation, changes in secondary traits

The effect of alleles in one locus can affect more traits. This implicates, that if changes by selection occur in one trait other traits will also be changed, namely those affected by the same genes. Often surprising results will occur. Therefore it is important that, in a selection program, the overall phenotype of the animals are maintained in an acceptable style to avoid unwanted secondary changes. The ideal is selection for one trait and proportional change in all the others, thereby maintaining a harmonic phenotype.

If some degree of relation exist between changes in two traits, they are said to be genetically correlated. The genetic correlation can be negative or positive, with a magnitude from -1 to +1. The genetic correlation has the symbol $r_{A1,A2}$, A1 and A2 being the breeding value of the two traits, 1 and 2.

An example from beef cattle breeding. Selection for fast growth is a common goal of selection. By doing so it is hard to avoid a change in the final adult weight. At the same time, the selection will also result in larger calves being born, which can cause difficulties during calving if the ratio between the weight of the calf and the cow is changed unfavorably.

Figure 8.6 shows a two-dimensional distribution with negative genetic correlation of 0.50 between the traits X and Y. The selection limit (the line) is laid so both traits have the same weight. It can be difficult to change both traits in a positive direction due to the negative genetic correlation. It is very easy, on the other hand, to get good results from the selection if both traits are genetically correlated positively, with respect to the direction of selection.

When selecting for more than one trait, one of the following selection criterions can be used:

- Tandem selection
- Independent selection
- Index selection

Tandem selection occurs by selection of one trait at the time.
By independent selection the selection limits for each trait are placed independently.
By index selection the traits are weighted in relation to their economic significance.

It is difficult to estimate the genetic correlation, as it can vary a great deal from one population to the other, it also varies to a greater extend over time than the heritability. Here is, non the less, a formula for estimating the realized genetic correlation. When the trait x is being selected for and trait y are changing at the same time, the genetic correlation can be estimated as:

Figure 8.6
shows a two-dimensional distribution with a negative correlation of -.5. The line indicates a selection limit when both characters have the same weight.

Index selection

$$r_{Ax,Ay} = (delta\ G_y/sigma_{Ay})/(delta\ G_x/sigma_{Ax})$$

The genetic correlation is estimated as the ratio between the standardized selection responses, when only trait x is selected for, and trait y follows passively.

After selection through many generations the selected trait and 'fitness' have a tendency to become genetically negative correlated. As a result of this, and the passing of time, the trait has to be put under selection pressure in order just to keep it on the reached level. If selection is abandoned the population will slowly return to the starting point with optimal fitness.

To avoid this negative effect some environmental changes might keep the population in balance on a level that differs from the starting point, see Figure 8.7.

When significant changes occur in both the environmental and the genetic background, the interaction between genetics and environment become an important factor. Interaction is reflected in the negative correlation to fitness.

Selection experiments to illustrate genetic correlation. Figure 8.8 shows three parallel selection experiments with broilers, they have been carried out by the Australian, Pym. The results were published in 1982. The experiment was carried out through 10 generations and the animals were tested from when they were 5 weeks old, till they reached 9 weeks. In the three experiments there were selected for respectively **Weight gain (W), Food consumption (F) and Food conversion rate (E)**. The curves for the three characters have the symbols W, F and E. The direct selection response was measured at the same time and also how the other non-selected characters have changed due to genetic correlation.

Figure 8.7

shows that if a trait changes dramatically, it is often necessary to change the environment too in order to maintain maximum effect of the selection by utilizing the interaction between genes and environment.

Furthermore it is clear that the selection has resulted in quite different phenotypes, depending on which trait the selection was for.
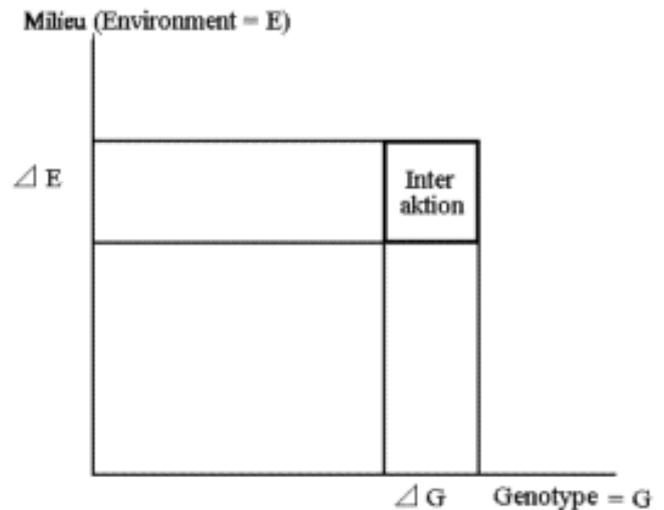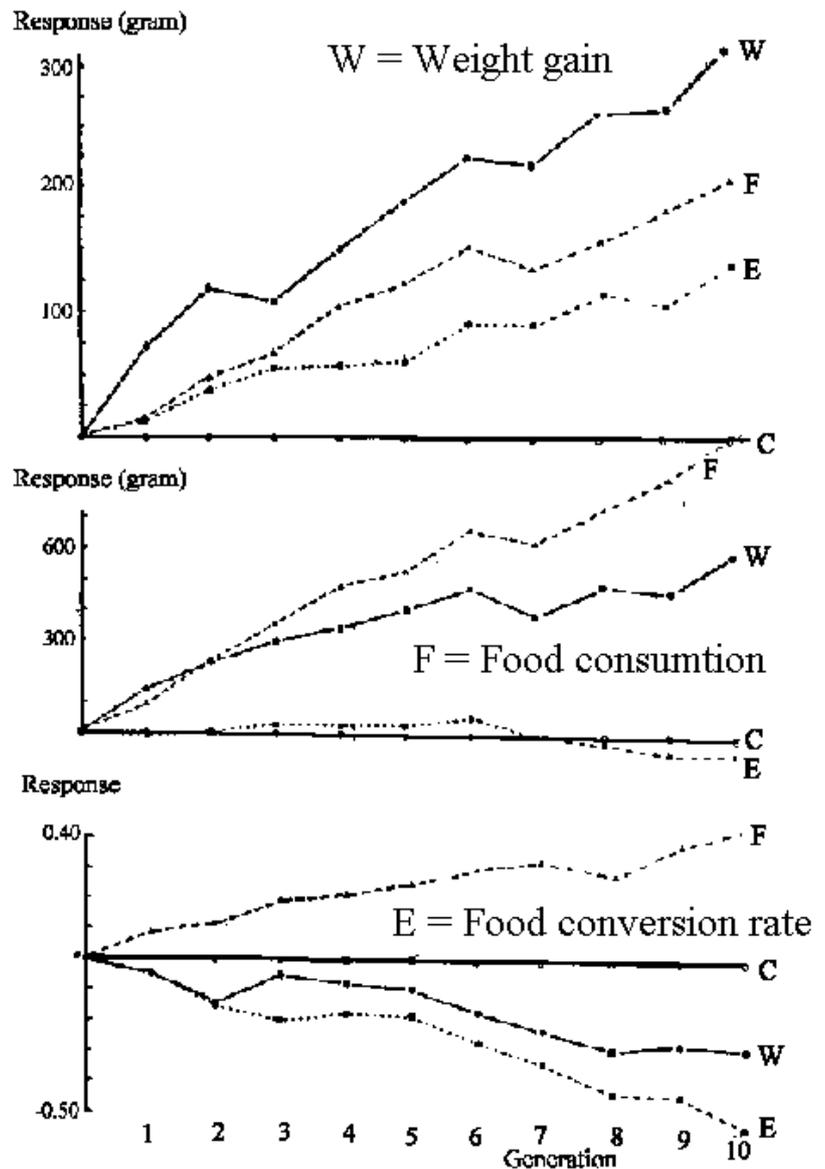
Figure 8,8
The Australian, Pym, have carried out three parallel selection experiments with broiler. The direct response is indicated in the title of each graph. The upper 'weight gain', the middle 'food consumption' and the lower 'food conversion rate'. All other curves are correlated responses.

Animals selected for growth (upper graph W) also has a larger intake of food (middle graph W) and a better efficiency (lower graph W) with a lower food consumption per kg weight gain.
Animals selected for food consumption (middle graph F) also has a bigger weight gain (upper graph F) and poorer efficiency (lower graph F) which mean higher food consumption per kg weight gain.
Animals selected for efficiency (lower graph E) also has a larger daily gain (upper graph E) and a stable food intake (middle graph E).

Video clips of broilers selected for daily gain with correlated leg problems, (demands installation of Real Audio)
The video clips are from a broadcast from DR 'Journalen' and is recorded at Denmark Jordbrugsforskning, Foulum, where prof Poul Sørensen has provided the basic information.
Video clip of broilers selected for daily gain and a corresponding layer chicken. At five weeks the white broiler weighs around 2 kg, which is about four times the weight of the brown layer chicken
Video clip of broilers selected for daily gain with correlated effects that produce leg problems. The first shown broiler's leg problems are so severe, that it is unable to get food and water, thus it ought to be euthanised. The second broiler walks badly, and keeps its balance by putting its feet wide apart.

Auto selection for the trait, litter sizes - balance with fitness.
In every generation an auto selection for higher litter size will occur, as more offspring are born in large

63

litters than in small ones. The selection differential in pigs is about 1 pig per generation, see Figure 8.9. The trait litter size is normal distributed when observing the trait in the sows. But the curve for the trait born in litter size is skew to the right, as more piglets are born in large litters than in small ones.

The maternal effect is positive for the pigs born in small litters, while it is negative for pigs born in large litters. In the selection of gilts it is important to avoid individuals born in large litters. The negative maternal effect is especially pronounced in species, which become sexually mature before they are fully grown, which is the case for pigs.

It must be clear that all multipara animals need a mechanism to maintain an optimal litter size, even though a strong selection for larger litter sizes always is present.

Relapse after relaxed selection for traits, which are part of fitness.

If a trait has been changed by selection, the trait has a tendency to return to the starting point if the selection is discontinued. The domestic swine, for instance, has a litter size around 10, whereas their origins, the wild boars, only have a litter size of 5.

Therefore it is to be expected, that if the domestic swine returns to nature, their litter size would adapt itself to the new situation where less resources are available. The same can be expected considering dairy cows, egg layers and broilers.

The highly specialized domestic animal species can only maintain their large production if housing and high quality fodder is provided.

Figure 8.9
the trait litter size is normal distributed when observing the trait in the sows. The piglet's distribution it is skew to the right with a mean value of one additional piglet.

Sow population
ave. litter size
10 piglets

4   7   13   16

Pig population
ave. litter size
11 piglets
(born in)

$S = 1$
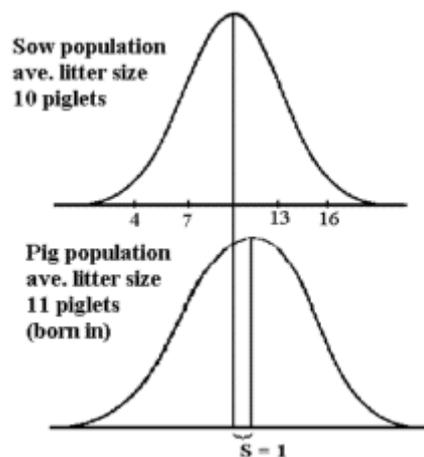
Chapter 9. Inbreeding, crossing and bred structure

9.1 Effect of inbreeding on individual and on population level

In section 4.4 on individual inbreeding, it was shown that inbreeding increases the chance of homozygosity in each locus. This is also true for homozygosity in general. Inbreeding causes an increase in the level of homozygosity, and a decrease in the level of heterozygosity. At 100 % inbreeding no heterozygosity is left.

What concerns an individual also concerns the entire population. Again we refer to the albumin example of dogs which, was given in section 2.4. By examining the total number of heterozygotes in all dog populations it becomes clear that 16 % units are missing, compared to the expected in a H-W population if we regard all examined dogs as one breed. The joint overall lack of heterozygotes makes it possible to estimate the inbreeding, within the individual population, to 33%, using the formula from section 4.4. The measurement of inbreeding in the various breeds can also be regarded as a measurement of differentiation between breeds. Inbreeding in a breed is caused by random loss of variation, whereby the breed loses parts of its original variation. Inbreeding drives the evolution, which is competition between populations on a higher level. Each population, if it is large enough, will acquire new mutations to compensate for the loss of original common variation.

It is known that one of the consequence of strong inbreeding is loss of vitality. A linear negative relationship between production traits and the degree of inbreeding can be predicted. Therefore it is important to ensure that the inbreeding is kept at the lowest level possible among the production animals. This can be done using pedigree information before mating, or by use of more robust systems, such as subdividing the animals into herds, keeping the females in the herd while adding new males from other herds. This system can ensure a low level of inbreeding, see Figure in section 9.3 for circular full sib mating giving minimal inbreeding. On the population level a herd can be regarded as a pair of full sibs.

The decrease in the level of the production traits is called **inbreeding depression**. It is at a magnitude of 0.4 units for an increase in the inbreeding of one unit. For reproduction traits or fitness traits the decrease is much higher, at about one unit. This cause the fact that it is extremely difficult to produce 100 % inbred domestic animals.

Figure 9.1 shows a linear decrease in average daily slaughter gain in swine. The slope of the line is 2.1 grams decrease for one unit increase in the degree of inbreeding (data from Christensen et al. Anim Prod. 58:298-300, 1994).

Figure 9.1.
Decrease in the average daily
slaughter gain in relation to
inbreeding in swine
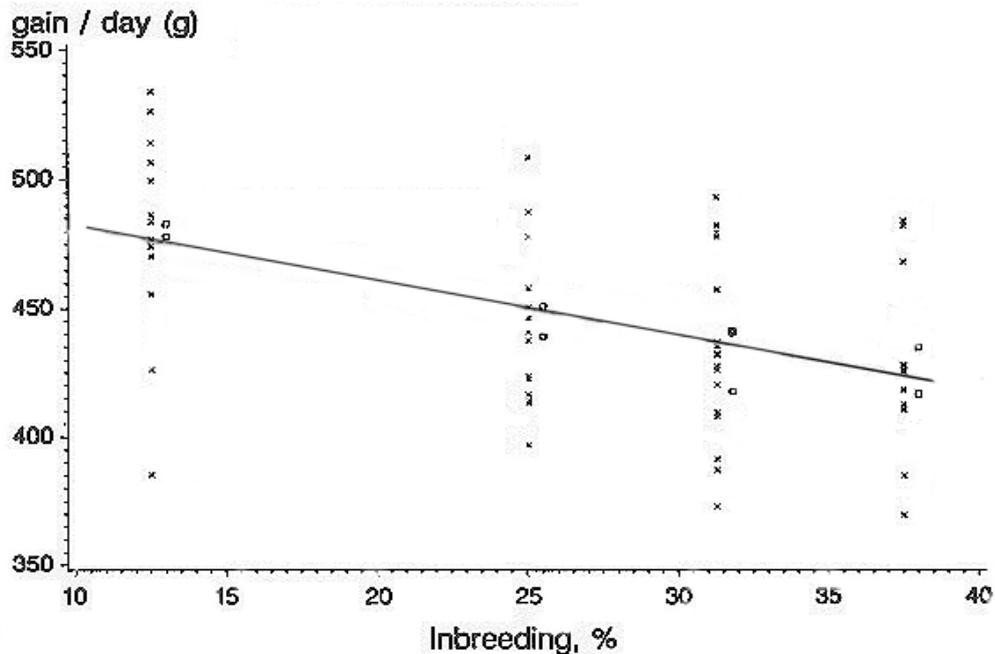
The linear decrease in production ability can be directly related to the linear decrease in the degree of heterozygosity, which is proportional to the increase in the degree of inbreeding. Correspondingly there is an increase in the degree of homozygosity, which is equal for both homozygote types. For every locus the decrease is proportional to $2Fpq$, and the corresponding increase is Fpq for each type, see table in section 4.4. The magnitude of the inbreeding depression depends on the magnitude of the dominance deviation, which is constant for a given trait. The inbreeding depression is generally largest in traits with low heritability.

9.2 Effect of crossbreeding

The effect of crossbreeding is the opposite of the effect of inbreeding. Crosses are more heterozygotic than individuals in a pure bred population. Contrary to inbreeding, where a linear negative relationship with fitness and production traits and the degree of inbreeding, it is impossible to predict anything about the effect of crossbreeding.

How is it possible to establish a sensible system of crossbreeding? The only possible method is 'the trial and error method'. Today a large part of the practical production of poultry, mink and pigs is based on production crosses, so the 'the trial and error method' is not as bad as it might sound. If a combination, showing a good crossbreeding effect or heterosis, is found, this combination can be repeated infinitely. Apparently, a certain heterosis does not depend on the level of pure breeding. **Pure breeding or out breeding** has not been defined earlier, but the words indicate that neither inbreeding nor crossbreeding occur, therefore breeding in a larger population of domestic animals is referred to as pure breeding or out breeding. This definition corresponds closely to the definition of a H-W population with less strict random mating requirements. **Hybrid vigour** is, as it says, a special vigour, which occurs as a consequence of crossbreeding. Another word used for a cross is a hybrid, which has then coined the term. Vigour is materialized in more robust animals with larger production capacity than pure breeds. **Heterosis** means the same as hybrid vigour. The word indicates that the effects are caused by increased heterozygosity.

Figure 9.2. Systems for crossbreeding.

Systems for crossbreeding. Figure 9.2 shows four breeds or lines, A, B, C and D. Crossbreeding can be carried out as two way, three way or four way crosses, backcrosses, or rotation crosses.

Two way crosses - a cross between two lines A and B.

Back crosses - the crossed animal, AB, mated with either line A or B.

Three way crosses - the crossed animal, CD, is mated with a third line A.

Four way crosses - the crossed animal, AB, is mated with the crosses CD.
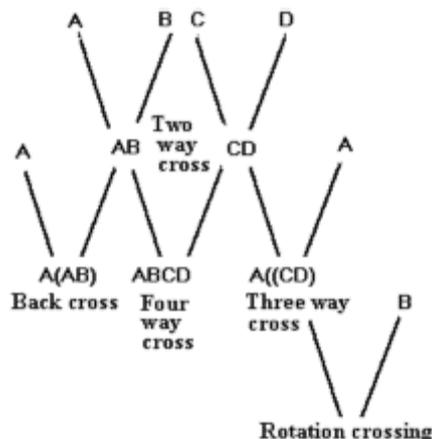
Rotation crossbreeding can be carried out with 3, 4 or 5 breeds or lines. In this case 4 is used, the next breed to be used is C and then D. Normally the females are crosses while the males are pure bred animals.
The two-way crossbreeding only gives heterosis in the offspring. To retain heterosis in the maternal traits, three way or four-way crossing has to be applied.
Systems for crossbreeding automatically ensures as low an inbreeding as possible in the production animals. Thus the advantage of crossing systems is that no special action has to be taken in order to maintain minimal inbreeding.
The traits, which give largest hybrid vigour, are traits with low heritability, which implies that hybrid vigour is mostly found in reproduction traits.

Figure 9.3 shows an example of hybrid vigour in mice on the number of young a female can produce in a lifetime. The most popular system for crossbreeding has been applied: two- and three- way crosses, and back crosses. (Data from Newman et al.
J.Anim. Sci. 61 358-365, 1985).

Figure 9.3. Effect of crossbreeding in mice.

As seen in Figure 9.3, the hybrid vigour is very high in this example as mice of pure lines only get half the number of young that the cross bred mice get.

The effects can be divided into either an effect on the foetus or an effect on the mothering abilities. The difference between the two lower curves shows the effect on the foetus, if it is a cross or not. The lower curve represents pure bred foetuses in pure bred mothers. The second lower curve represents a cross foetus in a pure breed mother. The difference between the two lower curves and the two upper ones shows the effect of a female being a cross or not. It is clear that most of the hybrid vigour is in the maternal part, and it concerns the ability to ovulate a large number of oocytes and hold them throughout the pregnancy.

CROSS-TYPE

□ PURELINE
○ TWO - WAY
△ BACKCROSS
✳ THREE - WAY
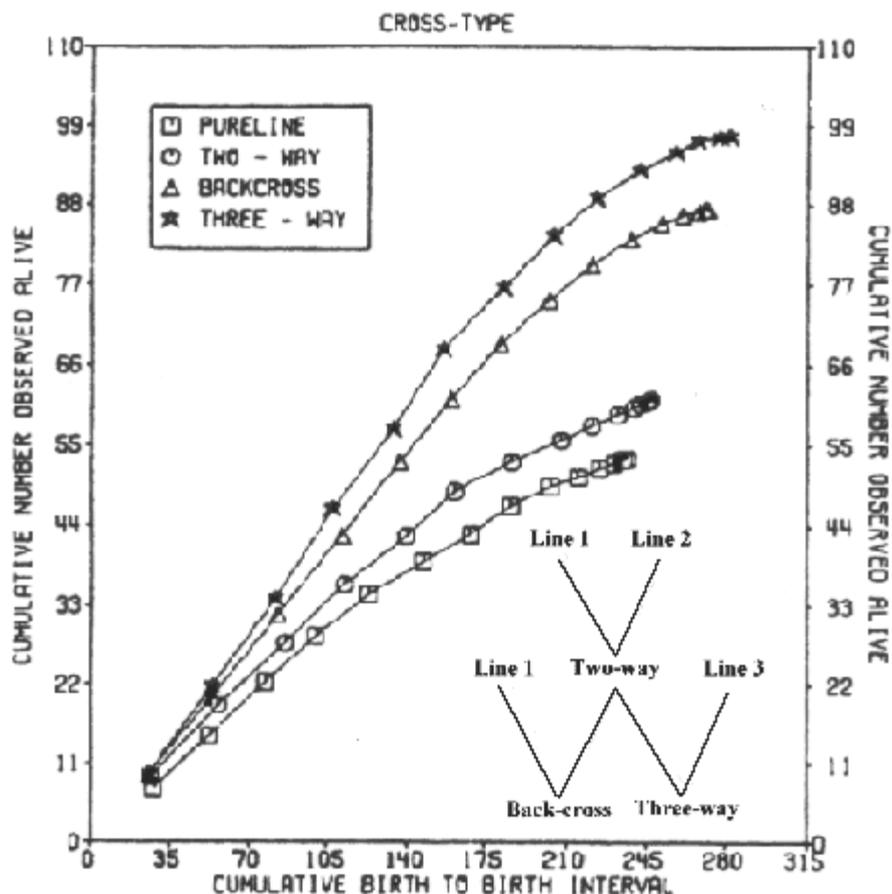
The example in Figure 9.3 shows a large amount of hybrid vigour. This is very rare, but in animal breeding it is common to see effects of up to 10 %.
Calculation of the hybrid vigour is done as follows: (average of the crosses minus average of the pure lines) in relation to the average of the pure lines.
Example of swine: Landrace Yorkshire crosses give an extra piglet per litter compared to the pure bred sows, which get an average litter size of 10.

Hybrid vigour (11-10)/10 = 0.10 or 10% for litter size in swine in the mentioned combination.

To carry out a cross production program it is necessary to maintain well defined lines, and that the animals are bred in sufficiently large numbers both to maintain them self and to produce a sufficient amount of animals to get into the crossbreeding program. The necessary surplus of animals for breeding is not always present. This however is not the case in the Danish dairy breeds. Largely all females are used for breeding, as the birth rate for the Red Danish and for the Holstein Frisian is only slightly above two. A ratio of 1.1 to 1.2 heifer calves have to born in order to replace their mothers, so the birth rate is just high enough to maintain the population. Rotation crossing though can be practised outside the pure breeding kernel. The hybrid vigour is significantly lower in rotation- than in three way crossbreeding.

An extra advantage using crossing is the uniformity, which is attained when at leas one of the parents is a pure bred animal. In a German investigation was found that the phenotypic variance in back fat thickness in swine was 30 per cent lower in L x Y crosses than in the corresponding pure bred animals (Lutaaya et al. 2001, 79:3002-07).

Figure 9.4. Two gene pairs and possible biochemical pathways. New combinations are not allways beneficial.

Negative recombination effect in F$_2$: In the shown crossing systems, at least one of the parent breeds was a pure bred. If an F$_2$ is produced, this is not the case. Production based on F$_2$-animals or crossbreeding between different crosses is normally a bad idea compared to backcrosses or three way crossing. The latter has the largest heterosis effect. In F$_2$s so-called **negative recombination effects can occur**. This is caused by combinations of genes, which did not exist in any of the original breeds. For instance if a F$_1$ is made by crossing aabb x AABB in F$_2$ animals of the type aaBB AAbb will occur, which

were not found in any of the original breeds. These types can be lethal, see Figure 9.4 for the biochemical pathways. The F$_2$ will furthermore be more heterogenous than all other crosstypes or pure breed animals.
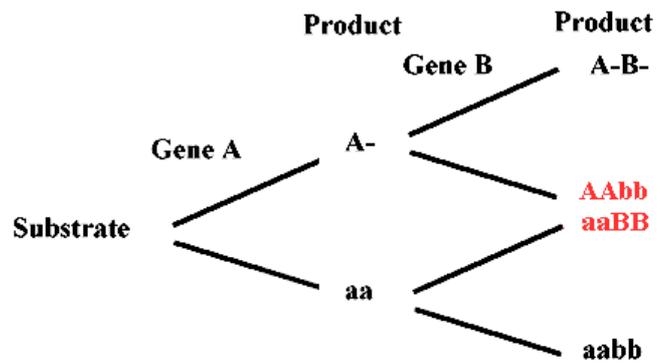
By **incrossing** it is possible to improve a breed by using a few superior animals from outside the breed. Normally animals from closely related breeds are used because they have the desired characters to save the best of the original breed. For this to succeed the selection must be light in the first generations of the incrossing. After incrossing some of the recombinant types are of higher interest and it can take several generations for them to occur

Figure 9.5 shows homologous chromosomes with four gene pairs which affect a quantitative character. Genes can be mixed so that it takes several generations for the desired combination, one chromosome having all the plus signs, to occur. See also section 2.5 for linkage disequilibrium.

Figure 9.5. A pair of homologous chromosomes from an original and an incrossed individual before recombination.
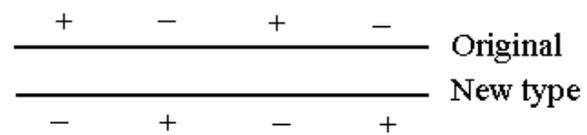
9.3 Minimum systems of inbreeding

In small populations it is often desirable that the general increase in inbreeding is as low as possible. One way to reduce inbreeding is to arrange the mating so that all animals have the same number of offspring, which are used as breeding animals. In principle every individual should be sire or dame of one male and one female offspring. If this is the case, the effective population size would be twice the actual number. This is due to the fact that random mating, as in a H-W population, in principle will have a left skewed binomial distribution of offspring per animal (a Poisson distribution) with the mean value of two. This is a distribution in which the most frequent number of offspring is 0, 1, 2, 3, 4 per individual. This means that a fair number of individuals in an H-W population does not get any offspring at all.

To ensure minimal inbreeding it is necessary to work systematically. The best and most simple systems use circular mating, as shown in Figure 9.6.
The figure shows a circular half- and a full sib system. The full sib system is easier to use in practice. The half sib system demands the mating of every individual with two different partners. In the full sib system both offspring are normally from the same litter. The figure shows that the half sib mating system is repeated in every generation, whereas the full sib system has a cycle of two generations.
The practical implementation of the full sib system is very simple. Mice for instance are kept in boxes. Place the boxes on a circle, and keep the female offspring in the box where it is born and put the male offspring into the box on the right.

Corresponding to the cyclic full sib mating system a population can be divided into four groups, numbered from one to four. The female offspring stay in the group, whereas all the male offspring are transferred to the group on the right. The animals can only be mated with individuals from within the same generation. In this system there is no need to keep track of every single individual. The degree of inbreeding would not be higher if a pedigree had been used for four generations, to avoid close relationship between mated animals.

Figure 9.6. Circular mating systems with minimum inbreeding.
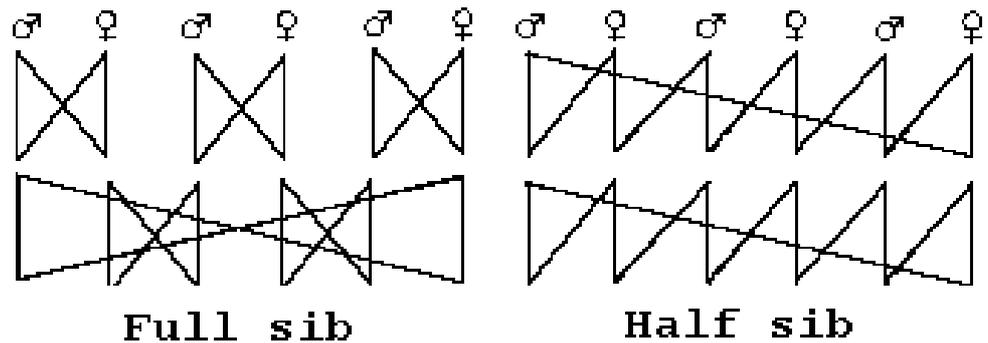
9.4 Inbred lines in laboratory animals

Within laboratory animals, as for instance mice, rat and guinea pig, it has been possible to obtain completely inbred lines. When using continued full sib mating it is possible in theory to reach 99.8 % inbreeding after 20 generations. If full sib or parent-offspring mating is continued the coefficient of inbreeding is determined by the inbreeding in two prior generations with the following formula

$F_n = (1 + 2*F_{(n-1)} + F_{(n-2)})/4$, where F_n is the inbreeding in generation n, see solution to exercise 4.2.

The totally inbred lines have the advantage of being completely uniform genetically. By using inbred animals (instead of out bred ones), in relation to a security test of a new drug, a lower number of test animals are needed due to a lower error variance. To maintain the uniformity the use of full sib mating is necessary. Furthermore, if several animals are used in the trial, it is necessary for the results that the multiplication ensures that the original parents are no more than five generations away. The reason for this rule is that every new individual carries a certain number of new mutations. These mutations are very significant due to the uniform inbred background.

It has not been possible to obtain absolute inbreeding within our common domestic animals. When the inbreeding has reach 60 to 70 %, the fertility has been very low because of the inbreeding depression, that further inbreeding has been impossible. The laboratory animals has passed that barrier and here many of the fully inbred lines have a high and stable fertility, even at 100 % inbreeding.

Every inbred line, that is maintained for commercial reasons, has been specialized for specific lines of research. An example is the diabetes mice which easily acquire diabetes, naturally they are used in diabetes research. Several transgene lines have also been produced in both rats and mice, each specialized for a certain purpose. Link to companies selling laboratory animals.

9.5 Population structure, breeding pyramid

The breeding of animals for production, as hen and swine, is almost entirely based on crosses. To maintain a steady production of crossbreds it is necessary to keep every breed, that are part of the final crosses, pure.

In such a system a breeding pyramid or kernel breeding including the following three steps are common.

- Breeding kernel

70

- Multiplication of parent animals
- Production

An example taken from the swine Danish production, see Figure 9.7. In Denmark more than 20 millions pigs are produced every year. The main part is 3-way crosses. The mother animals are mainly Landrace Yorkshire (LY) crosses. Most of them are mated with Duroc boars, a few with Hampshire breed boars.

In a lifetime a sow gets circa four litters with an average of over 10 piglets and little more than two litters per year. Which means that around half a million crossbred sows have to be recruited every year. They mainly stem from the multiplier herds. The breeding pyramid is shown in figure 9.7. The kernel consists of 50 privately owned herds (breeding centres) of pure breed. These centres breeds with eight thousand breeding sows, four breeds each with circa 2000 individuals. For each pig an index is calculated made up of the characters in the breeding goal. The males are mainly found at the AI stations (artificial insemination). The figure indicates, by use of arrows, the direction of the gene flows in the system. In the production herds around half of the piglets are offspring of AI boars. The other half are sired by natural mating. The boars in the herds are mainly bought from the breeding centres.

Figure 9.7. Breeding pyramid showing breeding system used in Danish swine production.
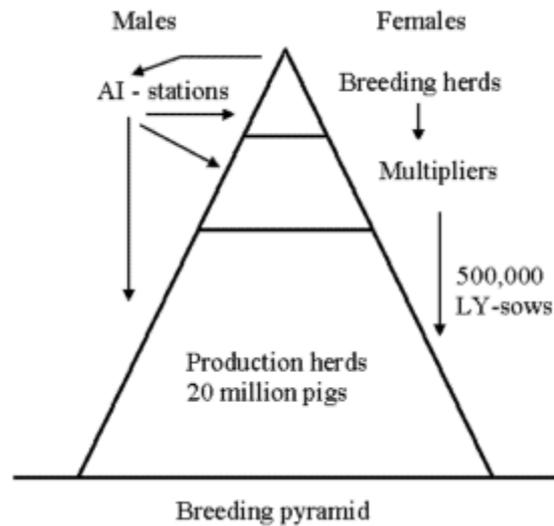
Chapter 10. Chromosomes and chromosome aberrations

10.1 Preparation of chromosomes

The chromosomes only become visible and suited for the classical chromosome analyses during the metaphase stage of the cell cycle. As the metaphase stage is very short, cells in strong growth are necessary in order to prepare chromosome slides for microscopic observation.

Lymphocyte culture is the most common method for obtaining many cells in the metaphase stage simultaneously. A mitotic rate of 1 % is common in good cultures. Normally the lymphocytes are inactive cells, therefore they have to be stimulated to divide. The most commonly used mitogene (mitose stimulating agent) is PHA (phytohaemagglutinin) which is a bean extract.

The following procedures are used for 'whole blood' culture and chromosome preparations:

- Use blood stabilized with heparin, which can be stored at room temperature (20 C^0) in up to five days before initiation of the cultivation.
- 0.3-0.5 ml full blood in 5 ml growth medium (RPMI) added 10% foetal calf serum and PHA is grown for 60 hours at 38 C^0
- the last two hours with added colcemid (then centrifuge and remove the supernatant)
- the cells are treated in half physiological salt solution (0,075M KCl) for 10 min. --- do
- the cells are treated in fixative (methanol acetic acid 3:1) added slowly while shaking, after 15 min. ------------------- do
- wash in new fixative three times, getting the cells in suspension every time --------------- do
- store in fixative at minus 20 C^0. The cells can now be stored for years before use
- drop cell suspension on slides and let it dry while blotting off the surplus fixative at the edges of the slides
- stain with Giemsa or another nucleus stain
- microscopic observation and photo

Chromosomes can also be made from fibroblast cell cultures, which can be established from an ear clip or subcutaneous muscle tissue. After a post mortem examination tissue from the lung or the kidney can also be used to establish a primary cell culture. The culture must be established within the first few day after the sample has been obtained. The cell culture is established in a Falcon tube by adding media and a few small lumps of tissue, which have been minced by a pair of scissors. Normally it has to grow for more than 10 days before a sufficient amount of cells for chromosome preparations have developed. The growth media are the same as mentioned under 'blood' culture, but here no PHA has to be added.

To identify the individual pairs of chromosome specific staining is needed. This is done by adding a thymidine analogue Bromo-deoxy-Uracil (BrdU) to the growing cells 6-7 hours before harvest, depending on growth rate. After staining with Acridine orange the BrdU incorporation should give faint cromatids in the late replicated areas (after addition of BrdU), whereas the earlier replicated areas give a strong bright colour. By means of this staining the inactive X-chromosomes in the normal female can be demonstrated, as the in section 2.2

10.2 Normal karyotypes in domestic animals

Metaphase chromosomes spread by means of the air drying method, make all the chromosomes of a cell stay in the same plane on the slide. By photo the cells that do not overlap are selected, so that the individual chromosomes can be identified. Additionally, the chromosomes have to be as long as possible. See for instance the blue fox metaphase chromosomes stained with Giemsa in section 10.3, which are fairly short.

The chromosomes can be set up pairwise, when individual pairs can be identified. A photo is taken and the chromosomes are cut out and arranged as shown in Figure 10.1. They are arranged according to internationally acknowledged enumeration systems. The enumaration systems are mostly ordered according to size and/or the position of the centromere. In each chromosome pair one of the chromosomes comes from the father and the other from the mother, these are called homologous chromosomes. Each species of domestic animals has specific chromosomes, regarding the number as well as the form. The number of chromosomes in mink is 30 and in dogs 78. Normally, animals with acrocentric chromosomes have the highest number of chromosomes. Therefore in the dog all, but the sex-chromosomes, are acrocentric. In the mink all the chromosomes are metacentric except for one pair. Examples of chromosomes from three species are shown in Figure 10.1.

Figure 10.1.
Cattle chromosomes, 2n=60,XY. Stain method
BrdU incorporation - Acridine Orange

. . Cat chromosomes, 2n=38,XX.
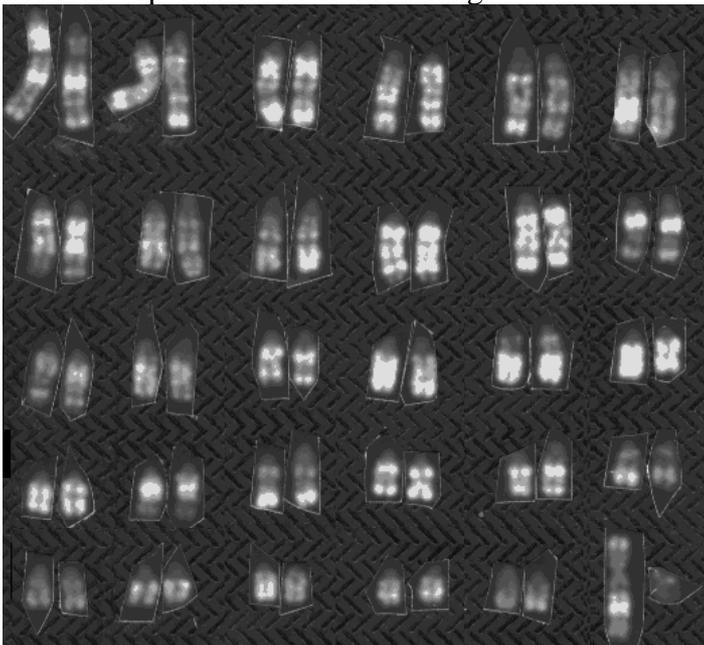
Swine chromosomes, 2n=38,XX.
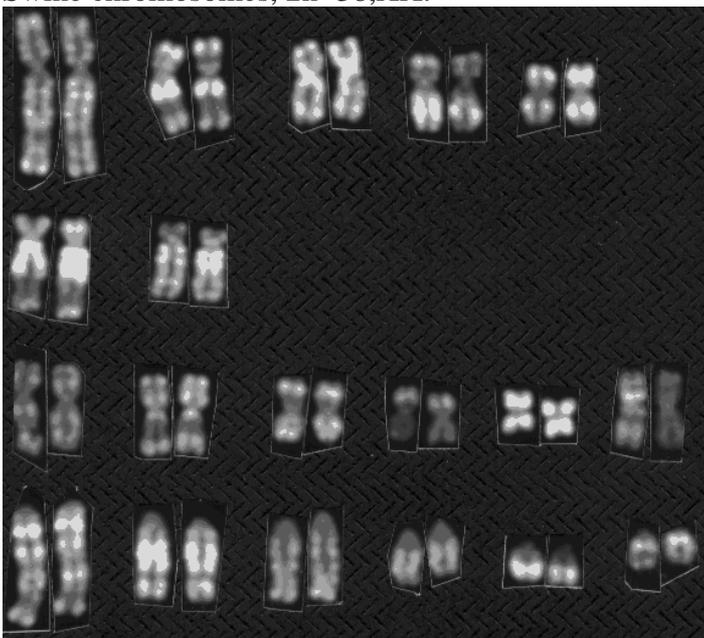
. .
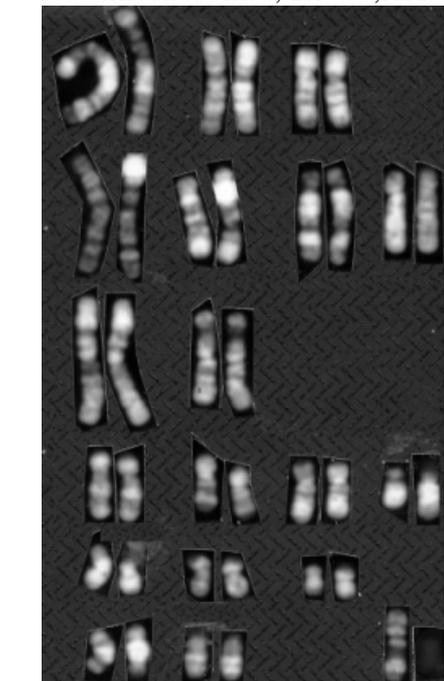
The Bovidae
Cattle chromosomes and chromosomes from sheep and goat are quite similar. Like cattle goat has 60

73

chromosomes, which are all nearly identical with those in the cattle, except for the sex-chromosomes X and Y. The X-chromosome in the goat is acrocentric (the cattle's X is sub-metacentric) and the Y-chromosome is much smaller than the cattle's. In sheep the same differences in the sex-chromosomes are found, but in addition there are three centromere fusions. The chromosomes, 1/3, 2/8 and 5/11, are fused in comparison with those in cattle and goat. Therefore, the sheep has only 54 chromosomes.

Regular fertile offspring can occur from the mating of a female goat and a male sheep. If the parents' sex is reversely combined the foetuses will die before term. The duration of the gestation period is almost the same for sheep and goat, around 148 days.

The Canidae

The two fox species found in Danish fur farms, the red and the blue fox, have respectively 34 and 50 chromosomes. The red fox has only metacentric chromosomes and one to five additional micro chromosomes. The significance of the micro chromosomes is unknown. The blue fox has two pairs of acrocentric chromosomes, the rest are metacentric.

After mating the two fox species non-fertile offspring occur in regular amounts. The duration of the gestation period is nearly identical for the two species, 58 days. A cross of the two species has at times been popular in the fur production.

Another canine species, the dog, has 78 chromosomes, which all are acrocentric, as mentioned earlier. So, compared to the Bovidae, the evolution of the chromosomes of the Canidae has been very fast. On the DNA level, in contrast, the Canidae species are very similar, see table in section 2.5, which deals with similarities of microsatellites in the three species.

Mink have 30 chromosomes, only, see next section or here. The mink chromosomes are somewhat similar to cat chromosomes, but are very different from those in foxes and dogs.

The Equidae

The horse has 64 chromosomes and the ass 62. The chromosomes in these two species are quite different. But non-fertile offspring can be produced from mating a horse mare and an ass stallion. This cross is called a mule, a very strong and enduring animal.

The Gallus

The chromosomes of hen are similar to those of other birds and reptiles. Six of the chromosome pairs are similar to those found in mammals, while the remaining are micro-chromosomes, 2n= 78. The sex-chromosomes are some of the largest ones. They are called ZZ in the cock and ZW in the hen. So among birds the female is the hetero gametic sex, which determine the sex of the offspring. Bird chromosomes can be seen here.

Mapping of Chromosomes

When working with chromosome mapping, the idea is to place a specific gene in a specific place on a chromosome. Therefore, for each species a specific standard with enumeration of each chromosome is made. For this nomenclature it has been decided that the shortest arm of the chromosome should be called the p-arm and the longest the q-arm. When the chromosomes are presented, the q-arm is always turned downwards. The enumaration of the bands on the arms begins at the centromere and continues towards the telomeres (the end of the chromosomes).

10.3 Chromosome aberrations in domestic animals

If an individual does not have a balanced set of chromosomes, i.e. two homologue chromosomes in each pair (and for the male an X and a Y), this will normally be visible in the phenotype, which then shows more or less deviation from normality. Animals with a non-balanced set of chromosomes will most often be sterile and have low vitality. Animals with a balanced set of chromosomes will generally be normal phenotypically. Chromosome deviations, in animals with a normal phenotype, are normally detected due to low fertility or

complete sterility. The karyotype of a bull with low fertility having a 1/8 translocation is shown in Figure 10.2, cf. Christensen, K., Agerholm, J.S. & Larsen, B. 1992a. *Dairy breed bull with complex chromosome translocation: Fertility and linkage studies*. Hereditas 117, 199-202.

Figure 10.2
Translocation between chromosomes 1 and 8 in cattle, 2n=60,XY. Staining method: BrdU incorporation - Acridine Orange. .
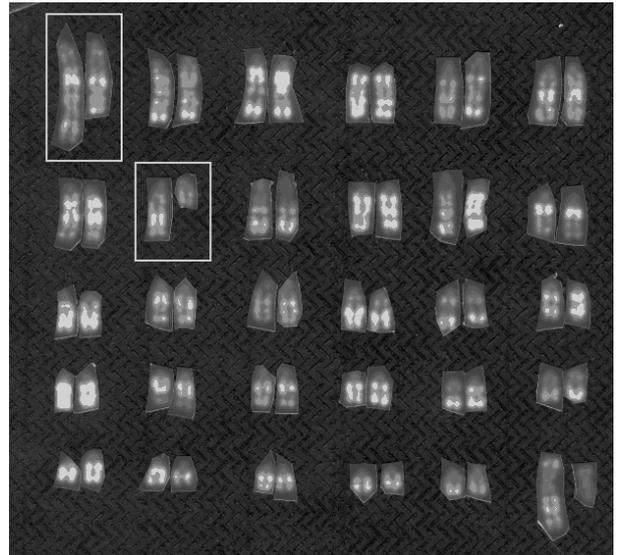
Other types of chromosome deviations worth mentioning are trisomies, the best known case being trisomy 21 in human. It occurs in about one in 500 new born babies.
The trisomies are very rare in animals, but they occasionally occur. Below is shown a trisomy 28 in cattle. The animal suffered from cleft palate and heart abnormalities, see Figure 10.3.
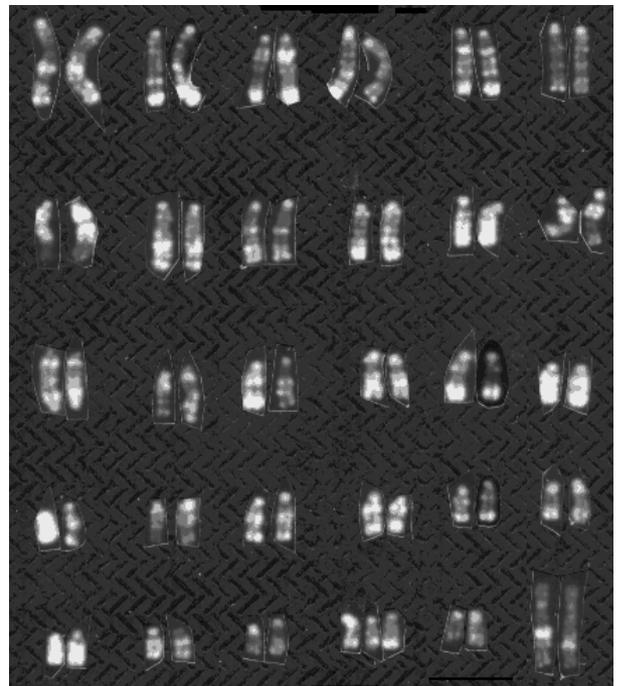
Figure 10.3. Trisomy 28 in a calf, live born but unable to survive, 2n=61,XX. Staining method: BrdU incorporation - Acridine Orange.

Normally the foetuses carrying trisomy 28 are aborted or die straight after birth.

In most domestic animals less severe chromosome errors occur. For instance fusion of two acrocentric chromosomes, called centric fusion.
The best known is the 1/29 centromere fusion of chromosome 1 and 29 in cattle. When such a fusion

occurs in heterozygote form, it causes a slightly lower fertility of about 10 %, when measured as return to the service rate.

A corresponding centromere fusion is found in the blue foxes. Here the effect is of the same magnitude concerning litter size, see table below where the chromosome number 49 corresponds to heterozygotes with respect to the centromere fusion. Data from Christensen et al. Hereditas 1982, 97:211-215.

```
Dam chromosome number  Number of litters    Litter size
      48                     16                 11.9
      49                     42                  9.8
      50                     17                 11.2

Sire chromosome number
      48                     10                 12.2
      49                     52                 10.1
      50                     13                 11.2
```

Metaphase chromosomes of the three types are shown in Figure 10.4. In the figure there is a circle around the chromosomes that participate in the centromere fusion.

Figure 10.4.
A centromere fusion of two chromosomes results in varying chromosome numbers between 48 and 50 in the blue fox

Blue fox chromosomes in 3 animals

50   49   48
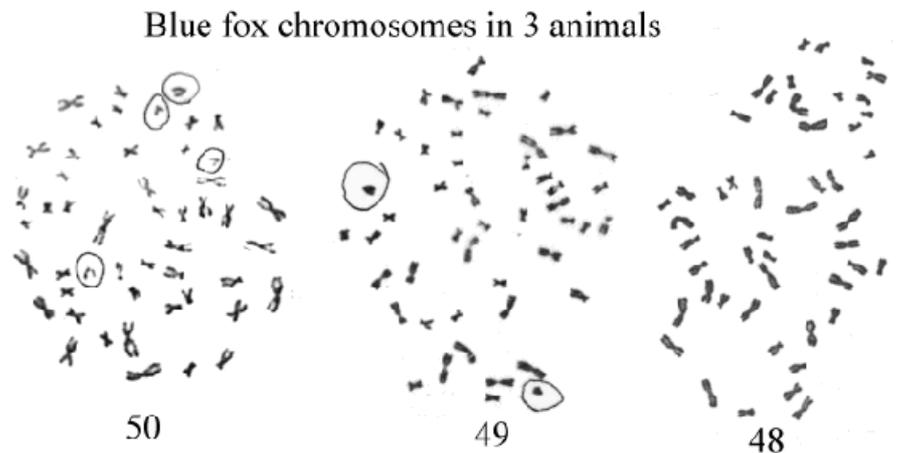
It is known that some mice carry up to seven sets of centromere fusions. When all fusions occur in heterozygote form these mice have a very low fertility, consistent with a reduction in fertility of about 10 % per centromere fusion in the heterozygote stage.

Figur 10.4a. Segregations of chromosomes from centromere fusion heterozygotes result in non disjunction and uniparantal disomy.

Imprinting: There exist centromere fusions involving all the 20 pairs of the acrocentric mouse chromosomes. If two animals heterozygotic for a centromere fusion are mated spontaneously there will occur offspring having received two chromosome of a pair from one parent and zero from the other, see Figure 10.4a. Such an individual has a balanced set of chromosomes, but is not necessarily normal. About half of the mouse chromosomes give rise to abnormalities when uniparental disomy occurs as there has to be one maternal and one paternal chromosome in a pair for normal development. This phenomenon is due to imprinting (which genes being on/off is depending if the chromosome comes from the sire or the dam). Imprinting is important to understand why cloning of adult cells give many difficulties. Incorrect maternal or paternal imprinting disturb the ratio between the development of the placenta and the foetus which often cause early embryonic morality and other developmental errors.

Non disjunction in centromeric fusions from heterozygotic parents

The ? are not always normal due to imprinting problems
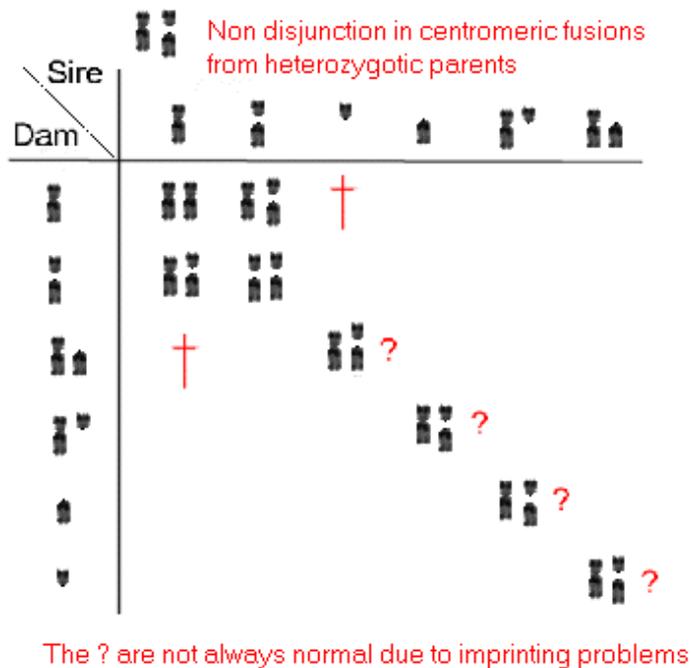
Free martins: Chromosome investigations can be used to identify animals with placental anastomoses, which often occur in cattle twins. When mixing the blood in the early foetal stages, a mixture of stem cells are established for the white and the red blood cells. The proportions are from 0 to 100 % of the right 'type'. If the mixing is too extensive the heifer in a mixed twin pair gets abnormal sexual organs and is infertile. The bull calf has normal fertility, but in paternity testing by means of a blood sample mistakes can occur, as this might show the genotype of the other twin. Therefore, when delivering blood samples for a paternity tests information should be given, if one of the involved animals have a twin.

Sex-chromosome errors in mares: Mares with XY sex-chromosomes are fairly common in some of the half-breeds. They are not fertile. There also occur a fair number of mares with XO or XXX sex-chromosome constellations. Figure 10.5 shows chromosomes of a mare with an abnormal X chromosome, which is always inactive. It has two q-arms. The picture is provided by I. Gustavsson, The Swedish Agricultural University.

Auli M�kin�, at the university of Helsinki has a good survey over [chromosome aberrations in the horse](#).

Figure 10.5. Horse chromosomes 64, XX. One of the X-chromosomes has two q-arms. It was always inactive.

Sister chromatid exchange (SCE)
Cells which are grown for two cycles, BrdU being added in the first cycle, can reveal SCE. SCE can identify individuals with unstable chromosomes or be utilized for a mutagenity test. The number of SCE is proportional to the doses of the mutagene. Below are shown mink chromosomes carrying SCE. The SCE chromosome is also called Harlequin chromosome. The chromosomes, without occurance of SCE, show no breaks in the chromosome, i.e. it has two unbroken strings, one light and one darkly stained. Where SCE

occurs the chromatids change their colour. For SCE to occur both strings of the DNA must have been broken to joint up with the opposite string.

Figure 10.6.
Mink chromosomes with 30 XY. Showing sister chromatin exchange (SCE).

10.4 Identification of chromosomes by means of chromosome paint

Figure 10.7 shows paint on swine chromosomes with a DNA probe, derived from a hamster cell line which contains the p-arm of swine chromosome 12. Two chromosome arms, which are identical to the two homologue chromosomes are painted.

Figure 10 7.
Paint on swine chromosomes with a probe containing the p-arm of chromosome 12
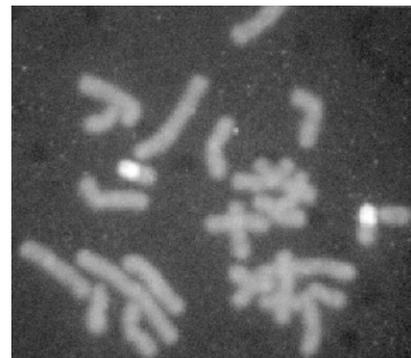
The paint method is summarized in the following. It is also called 'Fluorescence in situ hybridisation', FISH:

- Labeling of the probe with biotinylated nucleotides
- Melting of the probe and target-DNA (chromosome slides)
- Hybridisation of the probe and target-DNA under glass cover for several days
- Washing to get rid of non-bound probe
- A layer of avidin (binds strongly to biotin)
- Staining by means of fluorescence labeled antibodies against avidin
- Photographing by means of a fluorescence microscope

When paints with smaller pieces of DNA, for instance a 40-kb large, (inserted into a cosmid) very distinct results occur in the form of two dots, one on each of the two chromatids on the two homologous chromosomes, see here.

All human chromosomes are sorted so it is possible to buy pure DNA from every single chromosome. When each of them are labelled with different colours, the chromosome pairs also get a distinct colour of paint. This technique makes it simple to identify the chromosomes which are involved in complex translocation. Link to see paints of all human chromosomes.

By means of chromosome paint the homology between species can also be identified, see Figure 10.8 with mink chromosomes (left). The same chromosomes (right) are in situ hybridised (painted) with a human chromosome 14 probe. A table showing the homology between the chromosomes in four domestic species of animals can be seen here.

Figure 10 8.
Paint on mink chromosomes with a human chromosome 14 probe. Left are shown DAPI stained mink chromosomes. Right the identification by means of the paint which is on the p-arm of mink chromosome 13.
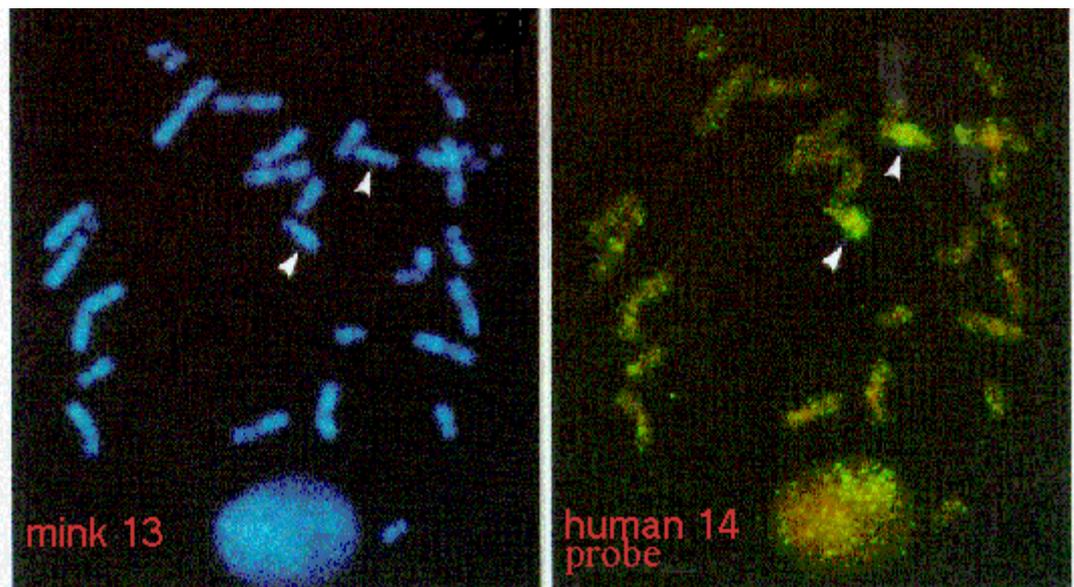
10.5 Chromosome aberrations identified by means of DNA-content in sperm cells

A normal male carrying a balanced translocation will produce both normal sperm cells and sperm cells in which the translocation gives rise to non-disjunction of the homolouge chromosomes which causes a production of sperm cells with deviating chromosome numbers.

Figure 10.9 shows histograms of flow-cytometry of sperm cells from two boars. A histogram represents an intensity of fluorescence, which is proportional to the DNA-content in each sperm cell. Each histogram contains an intensity of around 3000 sperm cells. The boar to the right (Number 1130) is normal. It produces sperm cells with a Y- and an X-chromosome with a ratio of 1:1. A sperm cell containing a Y-chromosome has ca. 4 % less DNA than a sperm cell containing an X-chromosome. This is reflected in the fact that the mean value of two distributions differs with 4 % units.

The boar to the left (Number 1523) carries a translocation between chromosomes 1 and 17. When the sperm cells are formed in this boar, some errors occur with respect to those two chromosomes, which are involved in the translocation. Chromosome 1 is very large, representing around 9 % of the DNA-content of the sperm cell. Therefore, the histogram of this boar has two additional symmetrical tops, one on each side. The small top to the right corresponds to sperm cells with two chromosome 1's, while the top to the left corresponds to cells with no chromosome 1's. The two extra tops only appear when they are companied by an X- or a Y-chromosome, respectively. An additional chromosome 17 or the lack of one is not directly visible. This chromosome is a small one containing about 2 % of the DNA-content of a sperm cell. Therefore non-disjunction of this chromosome only adds to the general variation of the two main distributions. It is also clear that the translocation boar contains more variation than the normal one.

The investigations are published by Jensen, P.O. et al. 1993. *Proceedings of 10th European Colloquium on Cytogenetics of Domestic Animals* pp 104-108.

Figure 10.9. Flow cytometry of boar sperm cells prestained with DAPI. The normal boar to the right has two tops, one for X- and another for Y-carrying cells. The boar with a translocation shows extra tops due to errors during the meiosis (non-disjunction).
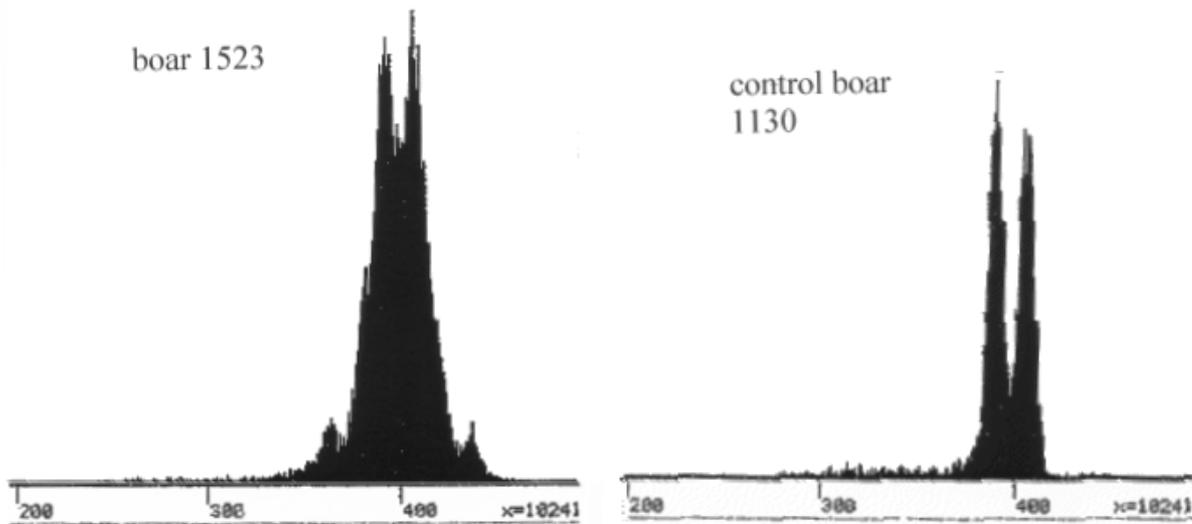
Naturally the interest in utilizing flow-cytometry in a preparative mode is big. Which means production of sperm cells carrying either Y- or X-chromosomes. All species of mammals have some difference in the DNA-content in respectively Y- and X-carrying sperm cells (both equally large) from 4 to 5 %. Technically it is feasible to work with a coefficient of variations (CV) at around 1 %, which is the case in the example shown above of the normal boar.

Until year 2000 the results, with respect to getting live offspring after insemination with sorted sperm cells, have only been done on an experimental basis. The flow-cytometry cannot be done without the sperm cells have been treated with papain and stained with DAPI or other flourocromes. The papain treatment is needed for the staining of the DNA to be uniform.

Chromosome specific DNA-probes to identify non-disjunction in sperm cells

Another method of identifying non-disjunction chromosome errors is the utilization of strong chromosome specific probes. After in situ hybridisation of the probe onto slides with fixed sperm cells, the number of signals per cell can be counted. The possible outcome is sperm cells containing 0, 1 (normal) or 2 of the chromosomes in question, depending on the number of positive hybridising signals. Investigations of this type on humans give an average non-disjunction rate per chromosome in sperm cells of .2 to .3 %, which equals a total of 5 to 8 % non-disjunction per sperm cell. If the rate of error in the female is of the same magnitude, the absolute maximum fertility rate will be at around 85 %, since almost all non-balanced gametes would die in the first trimester of foetal life.

Development of similar chromosome specific probes for domestic animals are under way. The few results already obtained confirm the results derived from humans. Examples of utilization of mink chromosome specific probes on mink sperm cells can be seen here.

Chapter 11. Genetics on hair and coat colour in mammals

11.1 Hair coat types in mammals

The hair of mammals can be characterized partly by the length and partly by the physical structure, for instance straight or curly. The ratio between the protective guard hair and the underfur also plays a role for its physical appearance. Finally the colour of the hair can give camouflage or in cases where individuals with strong colour dominates the flock or individuals outside the flock competing for dominance. Especially for water dwelling animals or animals hunting for fish the fur is of great significance for the regulation of heat. The fur of for example seal, otter and mink also has to be very resistant to last through the winter. The hare and foxes also have a warm fur, but the coat of these animals do not need to be nearly as strong as the before mentioned. Most mammals grow their winter fur in October-November and shed it again for the summer in April-May.

Some of the physical characteristics of the hair are under simple Mendelian inheritance, for instance the characteristics below:

Hair types and genotypes

```
Dominant                   Recessive
-----------------------------------------
Wire haired     W-         Straight haired ww
Short haired    L-         Long haired    ll
Guard  haired   M-         Mole           mm
Haired          Hr-        Hairless       hr/hr
-----------------------------------------
```

Figure 11.1 shows examples of the differently haired types.

Figure 11.1.

The wire-haired Airedale Terrier

Long haired type in Dachshund.

The fur from a mole mink and a normal Standard mink from year 1965.

Hairless mouse.

11.2 Coat colour types in mammals, colour genes

The colour of the hair is determined by the content of melanin grains. The black, brownish colours are produced by eumelanin and the reddish, yellow colours by pheomelanin. The latter can be dissolved in HCl. If the hair does not contain any melanin grains it becomes white, this is caused by air bubbles in the hair. This produces the white colour, just like air in or under the ice.

Melanin is formed by the amino acid tyrosine through a long row of biochemical reactions which all affect the coat colour. The melamine is produced in a cell type, which are called melanocytes. The number of and the hormonal influence on these cells can also give rise to variations in the colour.

Figure 11.2 shows a list of the best known colour genes.

Figure 11.2

Agouti locus (A locus)
Dominant Yellow A^y-
Dominant Black A-
Recessive Black and tan a^t-
Recessive Yellow aa

The Airedale terrier shown on the last page has the genotype Black and tan $a^t a^t$, black on the back and tan on the belly and extremities.

Brown locus (B locus)

Dominant Black B-
Recessive Brown bb

Brown genotype bb occurs in many dog breeds

Albino locus (C locus)

Dominant Colour C-
Recessive White cc

Albino genotype cc in the red-eyed Angora rabbit

Dilution locus (D locus)

Dominant Black D-
Recessive Silver dd

Silver genotype in swine

Extension locus (E locus)

Dominant Black E-
Recessive Yellow ee

Pink eyed dilution locus (P locus)
Dominant Black P-
Recessive Yellow pp
X-linked Orange locus (O locus)

Females:
Dominant Black oo
Heterozygote Black/Yellow Oo
Recessive Yellow OO

Males:
Dominant Black o
Recessive Yellow O

The heterozygote Oo has a mixed colour type due to random X-chromosome inactivation. The mixed colour can vary from a nearly perfect mixture to large spots of yellow or black depending on when the random inactivation takes place.

Somatic mutations in colour genes: In some domestic species, such as fox and cat, a dominant gene occurs, which reduces the pigmentation to a light cream colour, see Figure 11.3.

Figure 11.3.
Somatic back mutation in the Shadow gene for colour. The entire ear has normal colour on a Shadow blue fox, which is normally cream coloured all over the body.

The shadow mutation is very unstable as it easily mutates back to normal colour type. The somatic back mutation takes place under the foetal development while the melanocyte population is formed. In the foxes of the Shadow colour type a visible somatic back mutation occurs in about 2 % of a population. Figure 11.3 shows a back mutation in a

Shadow fox. Animals containing the Shadow mutation often have eyes with different colours, one brown and one blue.

11.3 The biochemical function of the colour genes

Figure 11.4 shows an illustration of a melanocyte. In the figure some of the functions of a melanocyte are pointed out. These functions are influenced by the colour genes mentioned in the last section.

MSH or the melanocyte-stimulating hormone is produced in the hypophysis. This hormone is important in the production of melanin. The MSH-receptor corresponds to the e-locus, whereas a blocker of this hormone corresponds to the a-locus.

The b and c locus code for 'enzymes' which regulate the restructuring of the amino acid tyrosine, tyrosinase. The restructuring of tyrosine is necessary for melanin to be formed. Locus d has to do with the dendrite-size.

Figure 11.4.
Shows the functions of the five best known colour genes in mammals. The coat colours are caused by the melanin grains produced by the melanocytes.

A comparison with the known alleles of the colour genes provides the possibility of an interpretation of their function. For instance can both a dominant and a recessive yellow exist in the agouti locus. The effect of the a and the e locus has to be strongly related, as they affect the same system, the MSH function.

The distribution of the melanocytes in the skin is also an important factor for the colour of the hair. An example of this is two gene pairs known from mice, the steel (sl or Mast cell growth factor=Mgf) and white spotting (the Kit gene), which is the Mgf receptor. These genes determine the number and the distribution of the melanocytes in the skin. They can also, as the latter gene name indicates, cause white spots. When fewer melanocytes occur in the skin the result will be a mixture of coloured and white hairs. This effect comes from the steel locus.

In mice up to 100 coat colour genes are known. They can all be categorized into one of the four classes given below:

Classifying genes which affect the coat colour:

- Genes affecting the biochemical path ways of melanine production
- Genes regulating the amount of melanin production
- Genes regulating the number and the distribution of the melanocytes
- Genes regulating the morphology of the melanocyte

More details and references concerning melanin production.

Modification of the colour can be done by means of classical selection, where the intensity of the colour is treated as a quantitative character. For instance, the colour of wild mink is dark brown. By selection for darker colour through 40 generations a completely black mink colour type has been created.

11.4 Colour genes in domestic animals

The inheritance of coat colour in the Danish swine breeds are regulated by the following three loci, see also Figure 11.5.

```
Dominant                        Recessive
-------------------------------------------------
White           I-              Coloured    ii
Black           E-              Red         ee
Uniformly       ii              White belt  i(be)i(be)
-------------------------------------------------
```

The Landrace (L) and the Yorkshire (Y) breeds are homozygotic for dominant white, II.
The Duroc (D) race is red, eeII.
The Hampshire (H) race is black with a white belt, EEi(be)i(be).
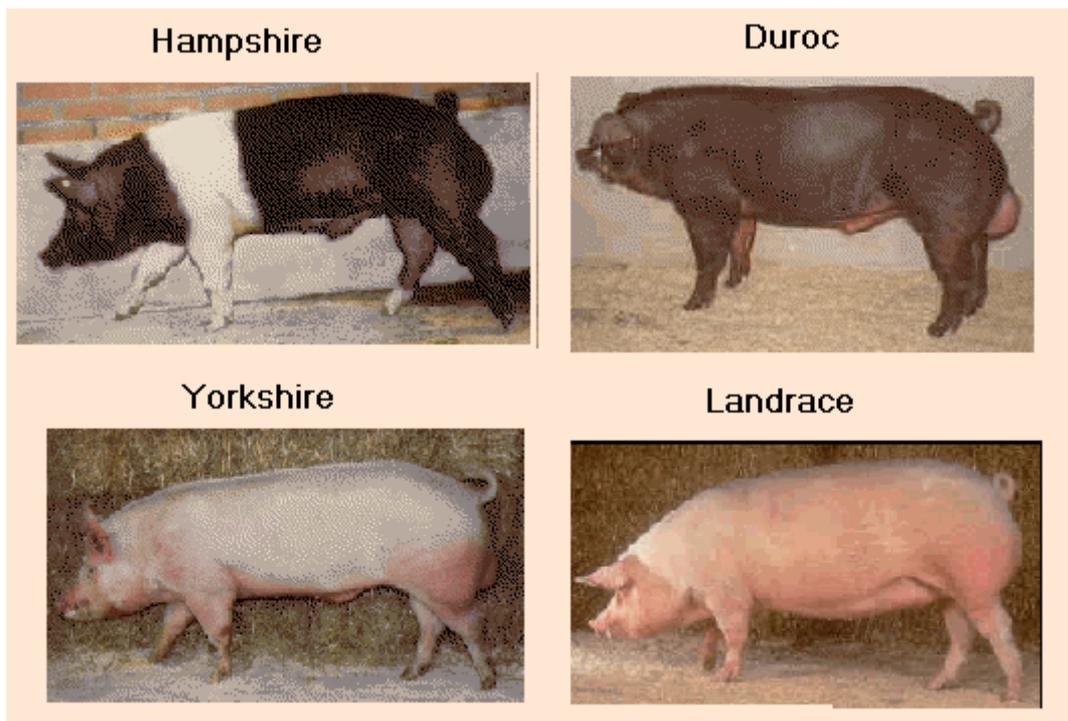
Figure 11.5
The Landrace and Yorkshire are homozygote for dominant white, the Duroc is red and the Hampshire is black with a white belt

Most of the pigs produced in Denmark have an LY mother and a coloured (D or HD) father. They are all white as they are heterozygote with respect to the genotype Ii. Marklund et al. 1998, *Genome Res.* 8:826-33, has shown that the dominant white in swine is caused by an allele in the kit gene, which corresponds to the mast cell's growth factor receptor. The extension locus code for the colour difference between the black and red pigs.

Details of the belt gene is given in Mamm Genome 1999 Dec;10:1132-6 af Giuffra E et al. it is an allele in the kit gene.

The inheritance of coat colour in the Danish cattle breeds are regulated by the following three loci: `

```
Dominant                    Recessive
----------------------------------------
Black         E-            Red      ee
Uniformly     S-            Spotted  ss
White head    S^H-           Uniformly SS
----------------------------------------
```

The Danish Holstein Frisian dairy breed (SDM) has the genotype EEss and The Danish Red dairy breed (RDM) has the genotype eeSS. Crosses between SDM and RDM become uniformly black. Some RDM have a different colour pattern, tiger stripes, which are cased by an allele in the e locus which dominates over red colour.

In some Danish beef cattle breeds, for instance Hereford, white head colour pattern occur, having dominant inheritance. It is inherited from a dominant allele in the same locus as spotted. It is known, that the gene for spotted is situated on chromosome 6, and is tightly linked with the kit gene and it has interaction with the kit gene.

Figure 11.6.
Danish Holstein Frisian with the genotype EEss and Danish Red with the genotype eeSS. The Hereford has the dominantly inherited white head.
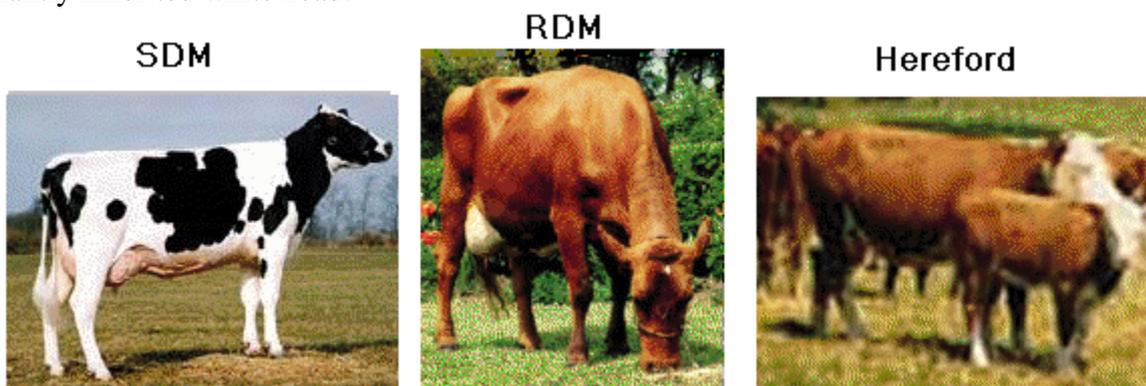
Klungland et al. 1995, *Mammalian Genome* 6:636-39, have shown that red colour in cattle is caused by the ee genotype in the extension locus.

Inheritance of coat colour in dogs, by [Helle Friis Proschowsky](#)
The inheritance of coat colour in dogs is very complex, therefore only the main types will be described here. In some breeds the coat colour is very constant, thus all individuals have the same coat colour. But each breed has its own special set of colour genes and some breeds have great variation. People with special interested in a particular breed can consult '*The inheritance of coat colours in dogs*' by Clarence C. Little, Howell Book House, 1984.

A short description of the best known colour loci and their significance for the coat colour in dogs:

Locus A: Determines the amount and the localization of dark and light pigment, both for the individual strands of hair and for the fur as a whole. The allele a^t, for instance, causes the black and tan colour of Rottweilers, Airedales and Gordon setters.
Locus B: Determines the amount of dark pigment and whether the colour should be dark (B-) or liver brown (bb).
Locus C: Determines the depth of the pigmentation. The allele c^a, gives albinos, and the chinchilla allele c^{ch} which makes the reddish and yellow colours become cream colour in for instance Cocker Spaniel. The ` chinchilla allele also gives rise to large variations in the depth of the pigmentation, in for instance Chesapeake bay retriever and Tibetan Terrier.

Locus D: Determines the intensity of the colour. The allele D gives an intense colour (black in Grand Danois), while the allele d gives a bluish dilution. Weimarans always carry the Recessive dd along with bb, thereby yielding the liver brown coat colour.

Locus E: Determines the pattern of the pigmentation in the coat. Alleles E^m, for instance gives the Leonberger a black mask and the allele e^b gives the Boxer and the Grand Danoi tiger strips (brindle), while the genotype ee produces the Labrador Retriever's yellow colour.

Figure 11.7 The black and tan colour in the Airedale, $a^t a^t$. Dilution colour in the Tibetan, dd, and yellow colour from the extension locus in the Labrador retriever, ee.

Locus G: alleles in this locus determine whether or not the colour fades with age.

Locus M: the dominant allele M gives merle colour (Collie, Shetland Sheepdog), while the Recessive allele m gives a uniform pigmentation. It is well known that two dogs carrying the merle colour should not be mated, as a double dose of the allele often causes damage to the eyes and the sense of hearing.

Locus S: determines whether or not the coat have spots. The dominant allele S gives a uniform pigmentation, though occasionally with white spots on the feet and/or the chest (Retriever, Boxer and New Foundlander). The three more Recessive s-alleles give varying amounts of white spots, the Clumber Spaniel and the Dalmatian being mostly white.

Inheritance of coat colour in the cat, by Nan Hampton, University of Texas at Austin, who also has an elaborated power point show with about 50 slides.

Below is shown a table of some of the more common genotypes and phenotypes of the domestic cat.

Common Genotypes and Phenotypes					
Homozygous		Homozygous		Heterozygous	
Genotype	Phenotype	Genotype	Phenotype	Genotype	Phenotype
AA	agouti	aa	non-agouti (black)	Aa	agouti
BB	black pigment	bb	brown pigment	Bb	black pigment
CC	full color	cbcbcscs	burmesesiamese	CcbCcscbcs	full colorfull colortonkanese
DD	dense pigment	dd	diluted pigment	Dd	dense pigment
ii	full pigment	II	pigment limited to tips of hairstrand, base of hair light gray or white, agouti especially affected by agouti=silver tabby with non-agouti=smoke with orange=cameo		
LL	short hair	ll	long hair	Ll	short hair

mm	long tail	MM	lethal on prenetal stage	Mm	short tail
oo or oY	non-orange (usually black)	OO or OY	orange	Oo	tortoiseshell=orange and black in females
ss	no white spots	SS	white spots > 1/2 body	Ss	white spots < 1/2 body
TT	mackerel stripe	TaTatbtb	Abyssinianblotched	TTaTtbTatb	Abyssinian, faint leg and tail stripes Mackerel stripes Abyssinian, faint leg and tail stripes
ww	not all white	WW	all white	Ww	all white

Inheritance of coat colour in mink

Mink's colour types have evolved through selection or mutations of colour genes. The wild genotype for colour, which has no mutations, is called Standard and ranges from dark brown (wild mink found in nature and the colour type *Wild mink*) to the strongly selected farm mink Black standard, which is totally black. This is due to selection for darker colour in the last 40 generations.

The mutated colour types in mink are classified according to numbers and types of genes responsible for the appearance of the colour. Until now it has not been possible to correlate the mink colour genes with the colour genes described in other mammals. The only exception is the albino locus. Hopefully this problem will be solved in the near future by means of DNA technology.

The Recessive and dominant colour genes: Single Recessive colours stem from homozygosity in mutated Recessive genes. Colour genes from two or more mutated colours, having originated from the same wild (standard) genes, are called allelic genes and will occupy the same chromosomal location (locus). The homozygous state of dominating genes is often lethal, resulting in a reduced litter size when heterozygotes are mated. Only the most commonly used combinations of the mutations are shown. A cross between non-allelic mutations results in the dark brown colour. This was the common type when the mutations originally were established.

Colour types, single	Mutation name	Mutation genotype	Wild type, Standard
White	Albino, red eyeds	cc	CC
	Hedlund White, deaf	hh	HH
Grey	Aleutian, bluish	aa	AA
	Silver Blue, grey	pp	PP
	Steel Blue	p^sp^s	PP
Brown	Royal Pastel	bb	BB
	Moyl	mm	MM
	American Palomino	kk	KK
Bicoloured	Black Cross	Ss	ss
	95 % White	SS	ss
	Finn Jaguar	Zz	zz
Combinations			
Bicoloured	Pastel Cross	bb Ss	BBss

Bluish grey	Sapphire	*aa pp*	AAPP
	Blue Iris	*aa $p^s p^s$*	AAPP
Sand	Pearl	mmpp	MMPP
Light sand	Violet	aakkpp	AAKKPP

Combination types used as production crosses. The crossing of non-allelic Recessive genotypes is practiced to produce various dark brown shades (called *Demi buff*, *Glow* or *Wild mink*).

The following three crosses (F₁) are commonly used to produce the Glow type

```
Sapphire x Pastel
Pearl    x   "
Violet   x   "

      F₁    x   Wild mink
```

The *Pastel* is the female due to its high fertility. The F₁ female can be crossed with Wild mink. The F₁ female is very fertile and the offspring will maintain the Glow colour type as the *Wild mink* breed true for the Standard genes. The *Mahogany* colour type is produced by crossing *Standard black* and *Wild mink*. This gives a very dark brown mahogany colour type. The Mahogany F₁ can be used in further breeding with the same result, therefore a pure bred Mahogany line can be established.

References: Experience from farm practise.
Nes, N., et al. 1988. *Beautiful Fur Animals and their Colour Genetics*. Scientifur, 60 Langagervej, DK-2600 Glostrup.

Below is shown mink colour types from the book *Beautiful Fur Animals*, with permission from Outi Lohi.

Black Standard   Wild Mink

Sapphire   Silverblue   Pearl   Pastel Sapphire

Colour genes in the horse
For a detailed description of the horse colour genes reference can be made to 'The horse coat colour genetics' from the Model Horse Reference website. Horse Colours Genetics or from the Veterinary faculty in Davis, California Horse Colours

The most important horse colour genes with indication of back ground colour genes :

Agouti locus (A locus)   back ground CCEE

Dominant bay, A-		Recessive black aa	

Cremello locus (Cr locus) — **back ground aaee**

'Codominant' chestnut, CC		Heterozygot, CCr	'Recessive' light, CrCr

Extension locus (E locus) — **back ground aaCC**

Dominant black, E-		Recessive chestnut, ee

Grey locus (G locus) — **back ground AACCEE**

Dominant grey, G-		Recessive bay, gg

Roan locus (R locus) — **back ground AACCEE**

Heterozygot roan Rr		Recessive bay, rr	Dominant RR, die

Dominant white (W locus) — **back ground AACCEE**

Heterozygot white Ww		Recessiv bay, ww	Dominant WW, die

The colour genes work in concert, the types shown above are representative for a group, in the real world there is a continuum of colours due to modifying genes.

The genes in the agouti and extension locus works as earlier mentioned together, so if a horse should be black it shall have the genotype aaE-; an aaee horse has chestnut colour. There is also interaction between all other colour genes.

Roan and ekstension locus are linked.

Locke et al. (2001) Animal Genetics 32:340-343, have shown that the cream dilution, Cr, allele is not coded from the C locus, as earlier symbols have indicated. The gene is not homologous with the tyrosinase gene in other species.

Other species

The rabbit's coat colour genetics reference is to Jackie Carey's Rabbit Colour Genetics

Chapter 12. Estimating- and biotechnology and disease resistance

The present chapter gives a review of the technological conditions, which are relevant in animal breeding around year 2000. Fast technological changes might create entirely new ways of breeding animals.

12.1 Technology for breeding value estimation

In chapter 7 simple methods for calculation of the estimated breeding values were given. Only data with uniform relation were included. This is a very simple form of estimation of the breeding value and it demands very little computer power, as it is based on the calculation of a single weight factor. This weight factor is dependent on the heritability, the number of individuals in the group, their internal relations and how they are related to the candidate. With the introduction of modern computers, the basis for a better utilization of existing data for estimation of the breeding value has materialized. The more advanced methods are based on systems of linear equation, in which every observation is provided with its own equation. Thereby the creation of a system of equations for calculation of the weight factors for every observation is completed. The model includes all observations from related animal as well as observations from related to related.

An example of related to related animals: The mother of an offspring, whose father is being evaluated. If the mother of the offspring is present in the model of the father's estimated breeding value, the preconditions for random mating does not have to be too strict, since the father's breeding value has been adjusted for deviations of the dam's from the population mean value.

The implied condition, that all observations must occur in the same environment, can also be taken more lightly, since for instance herd average can be taken into account in the models. To do this the model implies that more than one family per herd must be represented and that some of the sires in the herd should be used in other herds as well.

The most important methods for estimation of breeding value:
1. Selection index (SI)
2. Best linear prediction (BLP)
3. Best linear unbiased prediction (BLUP)
4. Animal model (AM)
5. Genomic selection

The selection index was developed in the 1950's and was utilized before the computer age. The SI is at present mainly used for model calculation, as it is useful for the evaluation of the effect of multi trait selection. If a certain set of economic weight factors are used, each trait will get an expected delta G's. Before the computer age it was common to pre-correct data, as for instance for calving age or slaughter weight, since these traits had some biological variation. Pre-correction is still applied to some extent.

With the production of faster computers, it has been possible to develop models, which calculate estimated breeding value for all animals in a population. When this is the case, it is easier to use all information from all the individuals, as these will give information on relations to all the others. BLP does not include environmental factors, which makes is less useful. The solutions obtained from a few animals are identical to the solutions obtained by SI. The BLUP and AM can simultaneous estimate environmental effects and the correction for them.

In the last three methods the relationship matrix is utilized at varying degrees. The relationship matrix is arranged according to the tabular method given in section 4.4. The solution of that many equations cannot be done explicitly, which have been taught when to solve two equations with two unknowns. The solutions are first based on guesses and then on recalculation until the solutions remain constant. This method is called iterative. It is possible, by means of this method, to estimate breeding values of millions of animals simultaneously. At present the AM method is used in the both Danish dairy and swine breeding work.

Genomic selection by Thomas Mark
Genomic selection is a new technology in which breeding values are predicted from genome-wide markers

in the form of single nucleotide polymorphisms SNP. The genetic maps are based on SNP and they enable us to divide the entire genome into thousands of relatively small chromosome segments. Then the effects of each chromosome segment are estimated simultaneously. Finally, the genomic breeding value equals the sum of all estimated chromosome segment effects. The chromosome segment effects can be estimated for a group of animals (ie a reference population); and for any remaining animal, only a blood or tissue sample is needed to determine its genomic breeding value. The chromosome segment effects apply to all animals in the population in which they were estimated, because markers are in linkage disequilibrium with the causal gene that they bracket.

Genome-wide information allows accurate selection of young animals provided that phenotypes from sufficiently many reference animals are available. This means that genomic breeding values are especially beneficial when traditional selection is difficult such as when phenotypic recording is restricted by sex and age (e.g. very beneficial for dairy cattle). However, conservative use of young animals without phenotypic information (relating to self or progeny) is advised to avoid potential negative side effects related to unfavourable mutations, unfavourable selection pressure on non-recorded breeding goal traits and high rates of inbreeding.

12.2 The significance of artificial insemination for estimation of breeding values

Naturally artificial insemination (AI) has had great practical significance in the breeding of farm animals. By means of AI it is possible to utilize elite animals to a much greater extent than by natural mating. But AI really comes into its own and shows its significance as estimation of breeding values becomes independent of the ownership. Who owns the daughters of an elite bull? When using AI no single person can own all daughters of an AI bull, see also section 12.1 for rules of estimation of the breeding value. The most important factors when estimation breeding values are accuracy and impartiality.

The most important factors for estimation of breeding value
1. The estimated breeding values must be as accurate as possible
2. The estimated breeding values must be independent of ownership
3. The estimated breeding values of all candidates must be comparable

From point 1 it becomes clear that AI increases the size of the family and thereby also the accuracy of the estimated breeding value. In order to comply with the conditions in point 2 and 3 it was formerly common practice to ship the animal to a test station. This ensured that the results were independent of the ownership. At the same time the results were comparable, since they were obtained in the same environment. Admittedly, test stations do have great problems with point 1. It is both difficult and expensive to place a large number of animals on a test station. The biggest problem concerning the test stations is the definition of their environmental condition. Which environment is most appropriate to ensure that the selected animals are adapted to the future production system? This question has no proper answer.

AI can solve the problems concerning point 2 and 3. Point 1 is at the same time dealt with in the best possible way. Extensive AI and adequate recording can ensure that the results obtained from the production herds can be used in estimation of the breeding values. This means that the selection takes place in a place which closely resembles the present production environment. Thus the collection of field data is much more advantageous than the former test stations for swine and dairy. The use of open recording system is only possible when extensive AI is practised. Since the offspring of several different bulls are present in a herd, it is possible to utilize information from each animal as deviations from the average of the herd, instead of using the absolute data. This is quite important, as a bull will not necessarily get a fair evaluation, since it is not used evenly in low or high production herds.

Test stations are still necessary in cases where AI has not been fully developed and in cases where the owner can influence the results in one way or another. Test stations can also be of some value in the measuring of characteristics, which is impossible to record in the practical production.

For cattle and swine the AI is so well developed that almost all former test stations have been closed. For hen, mink and rainbow trout test stations is still needed for comparable results.

The reproduction in dog and horse breeding is almost similar to the way AI is practised, eventhough most of these matings are natural. Here a special problem occurs though, a tendency to mate high rank animals or low rank animals. This problem can be overcome by working with a full Animal Model. Random mating is required if a simple model is used.

Other conditions can ruin the estimated breeding value, for instance partiality, therefore a good or bad result should not influence the recordings. When hip dysplasia in a dog is scored at a local X-ray clinic, the owner might not want to pass a bad score on to the Kennel Club, whereas he would very much like to record a good score. It is difficult to compensate for such a bias in relation to the recordings, even when very sophisticated statistical methods are utilized.

12.3 Transgene and transgenic animals

This section only gives a very short introduction of the topic. Only the main themes are summarized.

Methods for gene transfere: Micro injection of DNA into the paternal pro nucleus right after fertilization of the oocyte. The success rate is low, a few per thousand. There is always the risk of insertional mutageneses, i.e. the introduced gene inserts itself into a functional part of another gene and thus ruins its functions.

Micro injection or other forms of gene transfere into embryonal stem cells or other types of pluripotent cells by use of homologous recombination. After selecting for the transgene cells, which is possible in a cell culture. Selection for the recombinated cells can be done by means of a linked which, is part of the construction. Cultivation methods for foetal stem cells are only fully developed for mice. It is still on a trial basis for all of our domestic animal species.

After transgenesis in ES, or other foetal cell, the transgenic cells are injected into a blastocyst. Then a chimeric animal is formed, some of these transgenic would form gametes containing the inserted gene construction.

Homologous recombination. The targeting of genes becomes possible by means of homologous recombination. This is of great significance to study the function of an unknown gene (with known DNA-sequence). By means of homologous recombination so called 'knock out' mice can be created with a destructed gene function. By investigating the offspring of the 'knock out' mice it is possible to identify the function of a gene, which is completely unknown. So the 'knock out' mice have become the modern test tubes for identification of the functions of a gene.

Gene constructions.

A transgene is normally composed of a promotor and a structural gene. The promotor decides when and where the expression of the gene occurs. If the gene is to be expressed in the mammary gland the casein promoter is often utilized, as the casein is one of the important milk proteins. The structural gene is either with or without introns. If the gene originates from a bacteria, it is always without introns. Genes originating from eukaryotee can be cDNA or genomic. The genomic genes are normally very large, which is a problem. The larger the construction the more difficult it is to make a functional transgenic animal.

Motives for transgenesis.

- The gene needs to have a large effect on the trait in question, for instance at least four times $sigma_A$
- The gene should add completely new metabolic abilities, for instance the forming of an essential amino acid.
- The gene should add resistance against a serious disease, which have no cure.

Other motives for transgenesis

Gene farming - production of drugs

Organ donation -xeno transplantation
Many medical companies have started the experimentation and production of drugs by means of transgenic animals. Normally, a single animal can produce sufficient drugs for the world market, so the topic is of little importance for this section. The same applies for organ donation, an example is the use of a swine heart for a human. In order for this technique to work it is necessary to knock out most of the strong ordinary antigenic systems in swine. This is difficult, but what is worse is that the xeno transplantation is dangerous. It is well known that modern pests in humans, AIDS etc., stem from the animal world. It is a well known fact that swine carry a lot of endogenous retro viruses. They could be a large potential hazard to mankind if extensive xeno transplantation is used.

To get an overview of the topic see the following chapter [The Science of Transgenic Animals for Food Production](#)

12.4 Utilization of DNA markers

Use of DNA markers within families
If segregation of a recessive disease occurs within a family and if information of the localization on the chromosome exists and if flanking markers is known, it is possible to identify carriers of the gene, even if the genetic code of the gene is not known.

Method: Figure 12.1 shows segregation of markers A and B, as well as a disease gene (*). The two diseased animals in the figure are full sibs. All the sibs and the parents have been examined for alleles in locus A and B. Among the offspring the two diseased ones have the genotypes AABB. Based on the offspring the haplotype of the parents can now be interfered. The gametes AB must be carriers of the disease gene due to linkage.

Figure 12.1.
Genealogical graph of gamete types of three linked loci. The gamete types are deduced from lab tests of locus A and B and the segregation of the offspring. The disease gene based on the two diseased offspring.
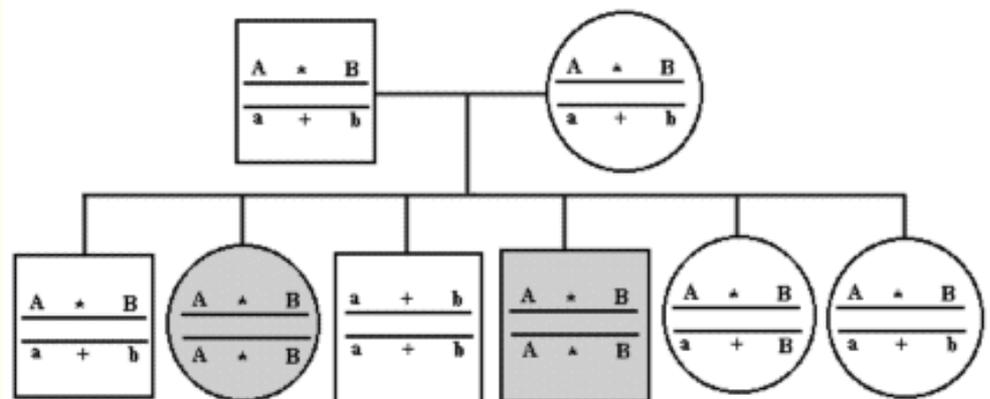
The fifth offspring has a recombination of the gene marker B and the disease gene, since is genotype AaBB and not affected.
In many cases only a single DNA marker exists, as not all domestic animal species can have that many markers. For the DNA marker to have any value it is necessary that the parents are heterozygotic.
Summary of the investigation: Animals with the genotype aabb do not carry the disease gene. Whereas all genotypes with at least one allele carrying a capital letter are expected to be carriers of the disease gene. By moving backwards in the pedigree, it is evident that all individuals with at least one allele with a capital letter are possible carriers of the disease gene.

DNA markers applied on population level
For a DNA marker to be applied on population level it has to be based on linkage disequilibrium and very close linkage with the disease gene. Strong linkage disequilibrium always exits if the mutation has occurred within a few generations back. A substantial amount of linkage disequilibrium occur even after 10-15 generations, if the distance to the marker is less than 1 cM, see section 2.5. Use of a marker on population

level would only indicate a possible carrier, which has a higher chance of being carrier than a randomly chosen animal. The final proof can only be established by segregation of diseased individuals.

To conclude this section on anonymous DNA markers it needs to be said, that use of a DNA marker for a disease gene is only second best. The goal should always be to find the mutation and to examine the DNA type to find the real culprit. Though the DNA markers are still to be used.

12.5 Detection of DNA markers for disease genes or QTL's

Detection of DNA markers in general
The size of the entire genome in mammals and birds is about 3000 centi Morgan (cM) or recombination units. This corresponds to that the largest chromosome is two to three hundred cM's long. The smallest ones are less than 100 cM's in length. There is a great deal of variation between the domestic animal species, the mink has only 15 chromosome pairs and the dog has 39. There are also some differences in the recombination rate between males and females, the females have on average 10 to 20 % higher frequency than the males. This is particularly true for the areas around the centromere.
Close linkage is easier to detect than loose linkage , and distances of more than 30 cM is in practise impossible to detect, since it requires the typing of a very large number of informative individuals (500-1000). On the other hand, if the distance is less than 10 cM linkage can be detected with less than 50 informative gametes.
For a full coverage of the entire genome, there has to be an informative marker for each 20 cM. All in all this gives around 150 DNA markers to be typed for complete coverage.
For the marker to be of any use it has to be informative, this will on average be the case for about 50-70 % of them. So to obtain a full set of markers in order to study a given family material, up to 300 evenly scattered markers have to be available. Only the parent animal has to be pre-screened for heterozygosity for all the 300 markers to get a useful set consisting of 150 informative markers to complete a full analysis of the entire family.

Useful family material for detection of linkage to a disease gene
The segregation of a new recessive disease normally occurs by inbreeding at the ratio of 1 to 3. The statistically most optimal cituation is for a dominantly inherited disease with a segregation of 1 to 1. For recessive diseases a 1 to 1 ratio can only be obtained by classifying offspring in a back cross.
To localize a recessive disease gene is needed more than 20 offspring, where 10 of which have the disease and the carriers parent animals. The analysis can start by marker 1 on chromosome 1 and performing a statistical test by continuing like that with a new marker until linkage is found. Linkage can be found in the first marker, but it can also be found in the last marker, number 150. On average only half of the markers (75) have to be applied before linkage is found. A statistically significant association would appear as follows:

```
                  diseased        not diseased
        Genotype ----------------------------------
        aa        |   10             3        |   13
        Not  aa   |   4             13        |   17
                 ----------------------------------
                  |   14            16        |   30
```

Here there are only of few recombinants, 3, with the genotype aa which not diseased, and 4 being diseased and not having the genotype aa.
To summarize: To identify a DNA marker for a disease gene animal material from a few related families should exist, comprising more than 20 offspring of which at least 10 should have the disease. For species where there exist a marker chip, used this of course as it is both faster and cheaper, you can also do with fewer animals because the chip has several markers for each centi Morgan, so it can be expected that there are markers which have no recombinants.
When linkage has been found, it is natural to continue using markers between the two markers providing the

linkage. The final goal will always be to identify the real disease gene. When linkage has been found, comparative studies can also be initiated. Candidate genes for the disease might be found by looking at the corresponding chromosome areas in other species, which are already known. An alternative to the classic marker analysis might be a careful study of the disease and thereby finding a candidate gene from another species. A candidate gene is a gene with a fair chance of causing the disease when comparing the aetiology of the disease. If one or more candidate genes exist, the analysis starts by typing these. If it is the right gene, complete association is found.

DNA markers applied for studies of quantitative trait loci (QTL) demand far more data material than it takes for finding a marker for a disease gene, if the gene are to be detected in a normal outbred population. QTL's correspond to the loci, which are discussed in the definition of breeding value in chapter 6. The most distinct problems in the detections of a QTL are that it is unknown whether or not a QTL segregates in a given family. If it segregates how big is the effect on a given trait? Another problem arises if a QTL is detected. How is it possible to discern the hypothesis that one or two genes cause the QTL effect? Planning of a QTL study has to be very carefully done in order to obtain any sensible results. A large amount of animals have to be typed in order to estimate a given QTL with sufficient precision. It is important that a number of characteristics are measured for the animals to be typed. Here all possible production traits can be mentioned, for example diseases and other easily observable traits. The optimal number of informative gene markers is around 150 with a distance of about 20 cM, as mentioned. An alternative to the study of QTL's in normal populations can be the study of F_2 individuals from special exotic crosses, as for instance between domestic swine and wild boar. The detected QTL's in such studies cannot be applied in connection with the normal breeding, but they can be used for pointing out candidate genes. The variation between F_2 animals from exotic crosses can be very large. Therefore, to find a certain number of QTL's the number of animals to be typed is smaller than in a normal outbred population.

Use of grand sire or sire design for detecting QTL's. The two classic designs, the grand sire and the sire designs are shown in Figure 12.2.

Figure 12.2. Showing the sire and the grand sire designs for detecting QTL's. With test of offspring for marker allele and phenotype measurements. The contrasts can be evaluated by classic statistical tests.
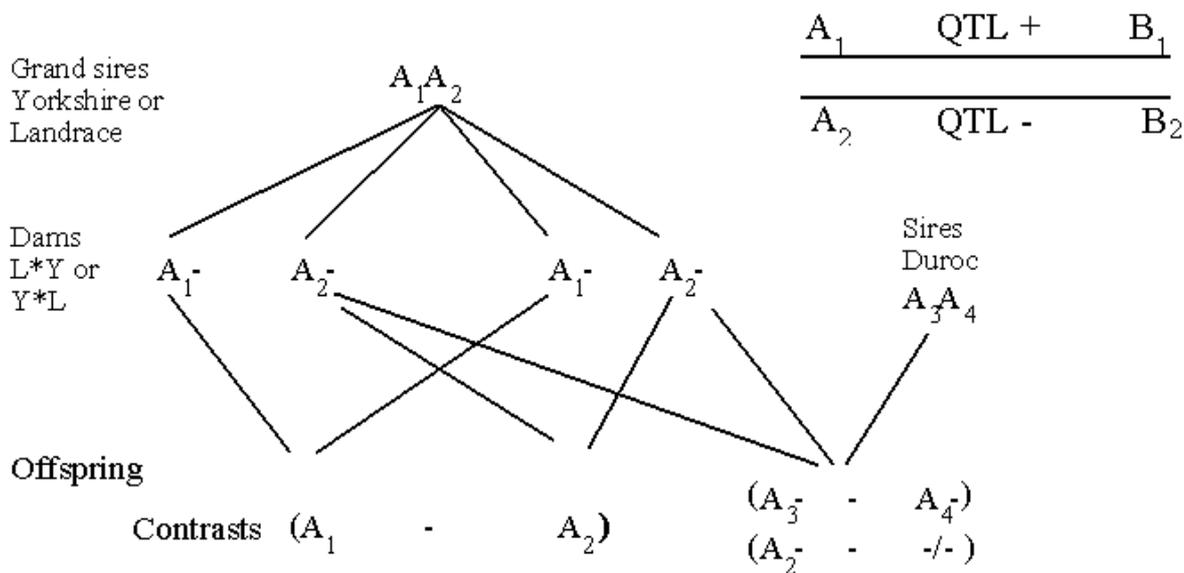
The grand sire is heterozygote with respect to the genotype A_1A_2. Half of his sons will receive the A_1 allele and the other half the A_2 allele. Now a contrast can be made between the average breeding value of those sons having received the A_1 and those having received the A_2 allele. In the grand sire design the classification of genotypes is only done in the sons, whereas the phenotype data derives from the grand daughters. This design has especially been used in the estimation of QTL's in dairy cattle.
In the sire design the genotype and the phenotype data come from the same animals. As can be seen from

Figure 12.2, more contrast can be estimated than in the grand sire design. In the sire design the major problem is that a large number of animals has to be typed. For instance, to detect a specific least statistically significant difference in disease frequency, when the average disease frequency is 10 %, the number given in the table below has to be typed. The binomial variance and a t-test contrast have been used for an approximation.

No with A_3	No with A_4	SD	least statistically significant difference
5000	5000	.006	1.8 % units
500	500	.019	5.7 % units
100	100	.042	12.6 % units

For 10 % units difference to be detected at least 200 animals have to be classified. As might be known, there are many causes for variation in the occurrence of disease and an average disease frequency of 10 % is fairly high. If the disease frequency is lower more animals is needed to detect a given QTL.

12.6 Results of experimental selection for disease resistance.

The literature displays many reports of selection against inherited diseases, but in most cases it concerns simple recessive traits, which, according to routine, are kept at an acceptable low level. Naturally, some problems arise in the cases where the selection is counteracted, which is the case when overdominance occurs. This is when the heterozygotes have a higher fitness than the homozygotes.

For diseases with polygenic disposition, only a few reports show successful experimental selection of the large domestic animals. The results, proving better resistance to diseases, have mainly been obtained by crossbreeding. But not only in theory, but also in practice, it has been possible to obtain significant results of selection against specific diseases. In the breeding of egg layers, the results have been just as clear as it has for experimental selections of laboratory animals.

Here is described an experiment carried out by Cole and Hutt (1973, *Anim. Breed.* Abstra. 41:103-118), see Figure 12.3.
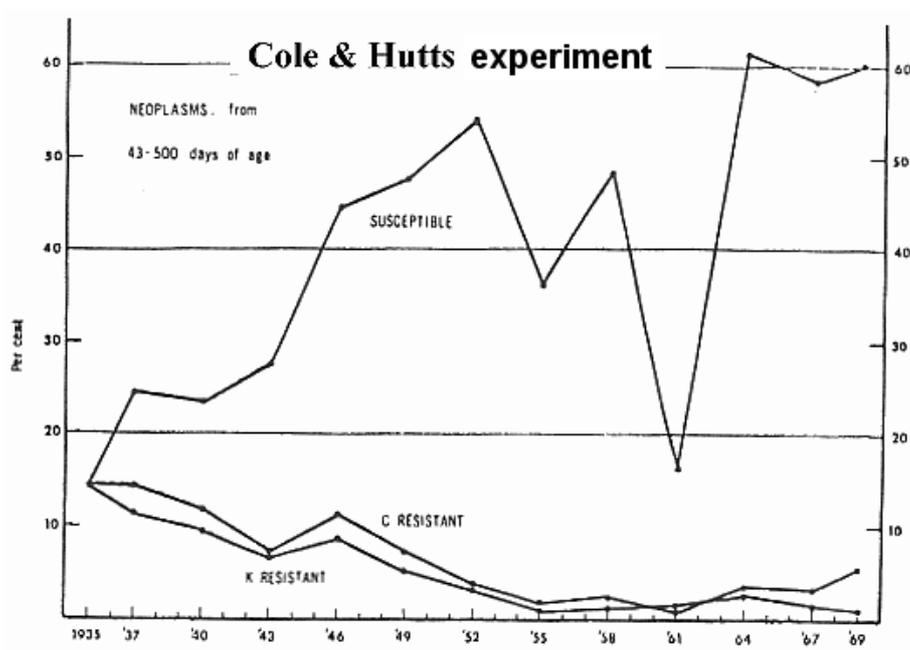
Figure 12.3.
Experimental selections of layers for resistance to neoplasma fatalities.

In 1935 two lines were established, C and K. The two lines respectively had 14.2 % and 11.3 % neoplasma fatalities due to leucosis including Marek's disease. In the period 1935-69, 34 generations were produced in each of the two lines.

The initial generation was not selected, but in the following generations the resistant breeding animals from the two lines were selected. In 1935 also a third line, S, was established, this one was based on individuals which were susceptible to leucosis (susceptible line). This line was maintained in the following generations by continued selection of the susceptible animals. The three lines were placed in the same production system. The environment was supposed to be infected by virus during the entire experimental period, except for the period 1958-61 (cf. the figure). The figure clearly shows that selection has been fairly effective concerning both points. At the end of the experiment (1967-69) the susceptible line had neoplasma fatalities of around 60%, while the two resistant lines, C and K, had neoplasma fatalities of respectively 3.7 % and 0.9 %. During the same period the total amounts of fatalities were reduced from respectively 51.5 % (C) and 44.4 % (K) to 12.7 % and 8.0 %. The fatalities in the period were counted in the period of 43 to 500 days after hatching. The mentioned results of the selection obtained even a fairly low estimated heritability, and were at the same time selected for several other traits. The good results could be explained by the fact that progeny test was used as a criterion of the selection. Furthermore Cole and Hutt argue that the same results could have been obtained in a significantly shorter time, if the strong progeny test had been used during the entire period.

As leukocytes are of great importance to the development of resistance, it is important to find out whether or not it is possible to select for a high or low content of leucocyte in the blood (WBC= white blood cell). Such an experiment has been carried out on mice by Chai 1975, J.Heredity 66:301-309. In the initial population (generation 0) the leukocyte counts were 6-8 x 10^3 per mm^3. After 22 generations of selection for HLC (high leucocyte count) and LLC (low leucocyte count) lines the counts were respectively 36-38 x 10^3 and 4-5 x 10^3. The examined control population had reached a level of around 10^3, which is close to the level in the initial population. The small difference is supposed to be caused by 'random genetic drifting'.
In relation to this experimental selection it ought to be mentioned, that the obtained response from the selection corresponds to h^2 at 0.20 for lymphocyte counts, which also has been estimated in cattle. If it is desirable, the lymphocyte count can easily be changed by selection.

Biozzi et al. (1972, J.Exp.Med. 135:1071-1094) made a selection experiment of mice for respectively high and low anti body response in relation to experimental immunization by means of sheep erythrocytes (SE), see figure 12.4. It appears from the figure, that the initial population had a titter with a mean value of around 1000 (ln 1000 = 6,9). After 20 generations of selection in each direction, the two populations had obtained titters of respectively 40 and 10.000. Furthermore, it was shown that the "high line" responded strongly to several other antigens, even though the selection was only done for responses to SE's. In the light of this it was easy to imagine that selection for general resistance to disease was possible, simply by selecting for high immunoglobulin concentrations in the serum. That this is in fact possible has been shown by Jensen & Christensen (1975, J.Anim.Sci. 40:392-396), who found that IgG 2 in dairy cattle had a heritability of 0.2.

Figure 12.4.
Experimental selections for high or low anti sheep erythrocytes titter in mice.

98
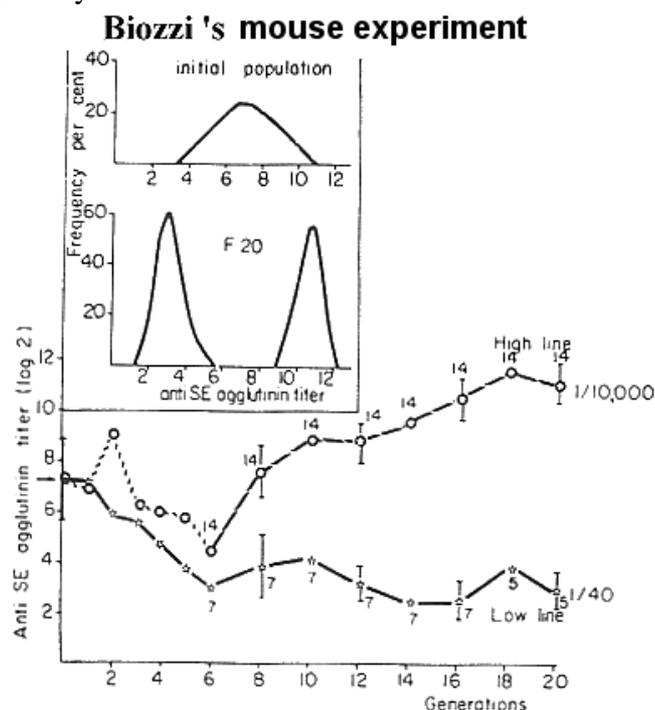
To test the generalization of Biozzi et al. (1972, J.Exp.Med. 135:1071-1094), concerning the results of selection for high and low anti body response in mice in relation to experimental immunization, a long series of experimental infections on the two lines of mice have been carried out. It is worth realizing that the fact that the line with a high general anti body response also tended to have a lower cellular defence based on macrophagical activity. This means that individuals with a strong cellular defence remove the anti-genes by means of their macrophages. Therefore, the anti-genes are removed before anti bodies can be produced against them. Nature has apparently found a compromise between the cellular and the humoral immune defence, see Figure 12.5.

But back to Biozzi's challenge experiments (Biozzi et al. Proc. World Cong. on Genet. Appl. Livestock Improvement, Madrid 4-8 oct 1982 Vol 5:150-163). They used challenge on more than 10 different pathogens among these were Salmonella thypimurium and Rabies virus. Their main conclusion can be seen in Figure 12.5. This figure shows that individuals, which have macrophage activity and immune response around average, are only slightly affected by most infections. Individuals with high immune response would be strongly affected by diseases, depending on macrophage activity as their main defence. The reverse is true for diseases depending on a high humoral response.

Figure 12.5. Challenge experiments with different micro organisms on mice selected for high or low anti body titter against sheep erythrocytes..
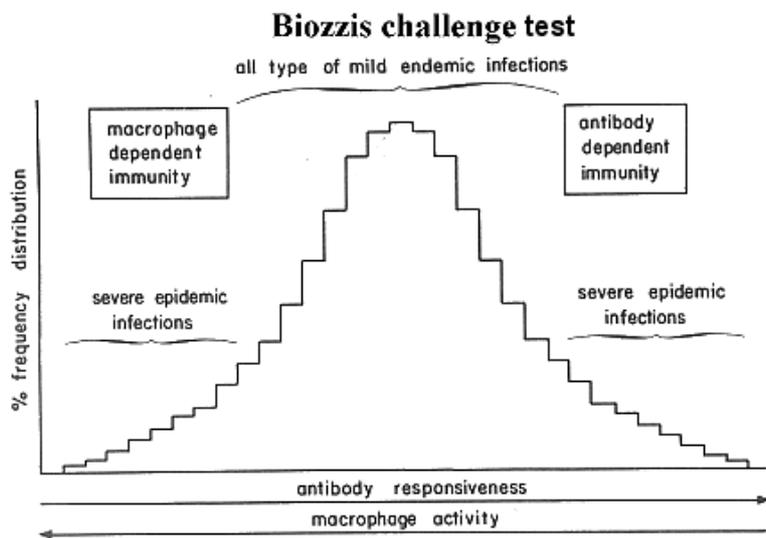
Conclusion:
The conclusions drawn from the old experiments are: It is possible to select for special resistance, but this type of selection will normally be overtaken by the possibilities of developing a vaccine. Today it is possible to vaccinate against leucosis among layers. Furthermore it must be concluded, that it is not possible to select for general resistance by selecting for high or low immunity components, as they act reversely. The natural level, with respect to infective agents, has to be balanced in the population history. Which means that pure breeding cannot be utilized to improve the general resistance. But what about crossbreeding?

By looking at Biozzi's graph from his challenge experiments it can be concluded, that the closer to the mean value an animal is the less severe disease it will have. The more extreme animals, on the other hand, would be exposed to the strongest attacks from disease. So a mean to obtain a low disease level in a population is to work with a uniform population of animals with an optimal mean immune response. Uniformity can be obtained in crossbred animals. Because of dominance conditions 2-ways, 3-ways and backcrosses are more uniform than pure bred animals. Therefore it should be possible to find combinations of crosses, which are optimal with respect to the exposure of a disease in a given production. F2 animals or 4-way crosses are more heterogene and therefore prone to acquire diseases. At present there are no practical experiments that can verify the presented postulates concerning different disease frequencies in different crossbred types. Therefore, experiments to elucidate these conditions are strongly recommended.

14. Genetic calculation applets and other programs

A java interpreter should be enabled in your browser to run the applets ! !
By using the applets a general feature is that you can place your input in the green fields. The yellow fields can also be used for input, often in the next round of calculations.
Be careful not to leave any blanks (not visible) in these fields, so if the applet will not run, test if the cursor stands right next to the number in all green or yellow fields you have touched.

Critical Values of the Chi-Square Distribution

df	.20	.10	.05	.02	.01	.001	.0001
1	1.64	2.71	3.84	5.41	6.63	10.83	15.14
2	3.22	4.61	5.99	7.82	9.21	13.82	18.42
3	4.64	6.25	7.81	9.84	11.34	16.27	21.11
4	5.99	7.78	9.49	11.67	13.28	18.47	23.51
5	7.29	9.24	11.07	13.39	15.09	20.51	25.74
6	8.56	10.64	12.59	15.03	16.81	22.46	27.86
7	9.80	12.02	14.07	16.62	18.48	24.32	29.88
8	11.03	13.36	15.51	18.17	20.09	26.12	31.83
9	12.24	14.68	16.92	19.68	21.67	27.88	33.72
10	13.44	15.99	18.31	21.16	23.21	29.59	35.56

Applets

2.2 Chi-square test for Hardy-Weinberg equilibrium

2.4 Calculation of mating type frequencies in a H-W population

2.5 Linkage, 3-point analyses from test cross

2.5 Many programs for statistical tests, collected by Lowry, USA

2.51 Linkage: Calculation of gamete and genotype frequencies

3.4 Selection: Change in gene- and genotype frequencies

3.5 Change in gene- and genotype frequencies: Effect of population size

4.5 Calculation of relationship and inbreeding coefficients by the tabular method

5.6 Calculation of Chi-square test for deviation from Mendelian ratios

5.7 Calculation of corrected segregation ratio according to Singles method

6.2 Calculation of mean, Genetic and Additive values and Dominance deviations

7.2 Estimating simple forms of breeding values

7.2 Estimating animal model breeding values

8.2 Calculating breeding values and selection response

8.3 Calculating selection index

2.2 Calculation of Chi-square test for deviation from Hardy-Weinberg equilibrium

- Put your observed numbers and click the **Calculate** button. if you see this you can't run the applet.

In all coloured cells with initial values you can put data (green fields, two alleles only, yellow fields for a third allele, and blue fields for a fourth allele)

The number of degrees of freedom for the chi-square is equal to one for two allele systems; and three for three allele systems; and six for four allele systems. The results are rounded (four decimals).

Example:
In a population was fund the following number of the three genotypes (obs), and there is calculated the gene frequencies for the alleles A and B and the corresponding expected (exp) under Hardy-Weinberg equilibrium.

```
---------
Genotype          AA               AB               BB       Total
Number, obs       36               47               23     = 106 = N
Frequency, exp    p²               2pq              q²     = 1,00
Number, exp       33,4             52,2             20,4   = 106
Deviation         2,6             -5.2              2,6
Chi-square        0.20             0,52             0,33   = 1,05
----------


Frequency  of A is calculated as  p = (2*36 + 47)/(2*106)  = 0,561
    do        B           do      q = (2*23 + 47)/(2*106)  = 0,439
                                                             -----
                                                             1.00
```

Try to put the observed numbers in the green fields of the applet and press the **Calculation** button.

Questions:
Calculate the gene frequency in the genetic system shown below and a Chi-square test for H-W equilibrium.

```
Genotype   Observed number
AA             31
AB             51
BB             23
CA             55
CB             71
CC             83
```

Is there statistically significant deviation from H-W equilibrium ?

In the Danish Landrace are found four hæmopexin alleles with codominat inheritance. In 1969 were typed 716 animals and the results of hæmopexin type are shown below:

```
        Hæmopexin type  Number
              0/0         2
              0/1        16
              1/1        75
              0/2        10
              1/2        67
              2/2        20
              0/3        21
              1/3       229
              2/3       116
              3/3       160
```

a) Calculate the gene frequencies for the four alleles.
b) Is the population in Hardy-Weinberg equilibrium ?

2.51 Calculation of Chi-square test for a 2 by 2, 3 by 3 or 2 by 3 table

In all cells with initial values you can put data (green fields), leave the zeros in the empty fields

The number of degrees of freedom for the Chi-square is equal to 1, 4 and 2 for the three possible combinations 2 by 2, 3 by 3 or 2 by 3 table.

if you see this you can't run the applet.

- Put your observed numbers, click the **Calculate** button.

Examples:
Use the applet to the test for genetic disequilibrium between two loci A and B. The number of observed gametes is as follows:

```
AB gametes 10
Ab gametes 20
aB gametes 20
ab gametes 10
```

The numbers are given in the initial values of the applet. Factor 1 is corresponding to locus A and Factor 2 is corresponding to locus B. Press the **Calculate** button to see the result.
The Chi-square are equal to 6.66 (df=1) which is larger then 3.84, this means that there is less than 5% chance that the genes at the two loci segregate independently.

Questions:
Calculate a Chi-square test for the association shown below.

```
    Genotype    Observed numbers
                Diseased   Healthy
       AA          15         27
       Aa          16         28
       aa          51         49
```

Is there statistically significant association between genotype and disease frequency ?
The two dominant genotypes can be merged into one with the following results:

```
    Genotype    Observed numbers
```

```
                Diseased  Healthy
        A-          31        55
        aa          51        49
```

What can justify to merge the to classes into one? Is there now statistically significant association between genotype and disease frequency ?

In a two gene system you have the following set of observed numbers:

```
     Genotype      BB        Bb        bb
              ---------------------------
        AA    |    57       140       101
        Aa    |    39       224       226
        aa    |     3        54       156
```

Are there statistically significant association between the segregating genes at the to loci ?
In case there is found association, what does this mean ?
Can you propose a chi-square test which give 6 degrees of freedom to test a hypothesis of linkage between the genes A and B ?

2.4 Calculation of mating type frequencies in H-W population

In the green field put the gene frequency for the dominant gene

if you see this you can't run the applet.

- and click the **Calculate** button.

Examples:
If a recessive disease gene occur with a frequency of .005, 99 percent of the diseases would be in families where both parents are normal healthy individuals.

Questions:
Which mating type produce the highest number of heterozygotic offspring when the gene frequency p = .8 or .4 in a random mating population.

2.5 Linkage, Calculation of gamete and genotype frequencies generation after generation

- Put your observed gamete frequencies or click the **initiate Mendel** button. If the mothers gamete frequencies sum to zero, the frequencies for the fathers is also applied for the mothers.
- Click on **Next generation** button calculate the gamete and genotype frequencies in next generation. if you see this you can't run the applet.

In all cells with initial values you can put data (green fields) and to get the results press the **Next generation** button. Table values are rounded to six decimals.

Example:
When you look at segregation of genes at two loci the gametes that can be formed are shown below and symbols are given for the observed and expected frequencies, the deviation between the two has the symbol D for disequlibrium.

```
 Gamete  Observed frequency   Expected frequency    Deviation
 -------------------------------------------------------------
  AB        r                   p(A)*p(B)              D
```

```
Ab      s               p(A)*q(b)               -D
aB      t               q(a)*p(B)               -D
ab      u               q(a)*q(b)               D
```

The disequilibrium gets the symbol D can be calculated from the table above, and can also be calculated as the half of the difference in the frequency of double heterozygotes in linkage- and repulsion phase:

```
D = r*u - t*s     (= [f(AB/ab) - f(Ab/aB)]/2) from the table above
```

The maximum disequilibrium occurs, when all double heterozygotes are in linkage phase f(AB/ab) or in repulsion phase f(Ab/aB). See the Figure below.
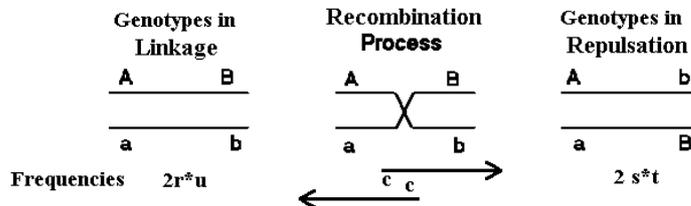
Are the recombination frequency c would D be reduced to D*(1-c) per generation, and for the n'th generation $D_n = D_0*(1-c)^n$.

If you apply the initial values r=0.1, s=0.2, t=0.3, u=0.4 and press the **Next generation** button, you will see calculated all the genotype frequencies and D=0.0199 rounded 0.02. If you continue using the **Next generation** button more times with c=0.5 you will see that D= approach 0 which means observed and expected gamete frequencies are equal i.e. r=0,12.

Questions:
Calculate the gamete frequency in the first 5 generations and the equilibrium gamete frequencies in the genetic system shown below where the recombination frequencies between the two loci A and B is c=0.2.

```
Gamete   Observed frequency
AB          0.21
Ab          0.49
aB          0.19
ab          0.11
```

What would be the frequencies of the two types of double heterozygotes in next generation?
How many generations will it take to reach equilibrium within 0.02 units ? Repeat the calculations for c=0.05.

In a F_3 after a cross between two pure lines aabb X AABB. What would be the gamete frequencies if c=.5 , .4 and .1 ?
Hints: Use the initiate mendel button and use the next generation button more times.

3.4 Selection: Change in gene- and genotype frequencies by selection

In all cells with initial values you can put data (green and/or yellow fields)
You can put either real observations (numbers) or relative genotype frequencies in the green fields. if you see this text you can't run this applets

Click the **Calculate** button. Put the fitness of the genotypes (relative or numbers) and you can run one generation at the time by pressing the button +**1 generation!**, or you can get a graph over gene frequency changes over 60 generations by pressing the **Run** button. The corresponding data you can extract over the clip board in the lower left window.
You can also start the process by putting your gene frequencies and there after take some selections rounds.

For practical reasons delta q is put to 0 for values less than .000001. Sometimes when delta q is less than .0001 there are problems with rounding and some browsers can not see the results proper.

Example:

If you have a genetic system for instance yellow coat colour in the Labrador Retriever, where the recessive (yellow) have the fitness 1-s by selection you get:

```
Genotype          EE              Ee              ee        Total
-----------------------------------------------------------------
Observed number  141              80              11       = 232
Frequency         p²              2pq             q²        = 1.00
Fitness           1               1               1-s
Proportion        p²              2pq             q²(1-s)   = 1-sq²
after selection
---------
```

After selection the gene frequency q' can be calculated by the gene counting method, q' is the gene frequency in the next generation, is calculated as half of the heteroygotes plus the surviving recessive relative to half of the proportion of all surviving genes, that is equal $1-sq^2$.

$$q' = (2pq/2 + q^2(1-s))/(1-sq^2)$$

To test the applet put the observed number and press the **Calculate** button. You then get q=0.2198, if you have s(aa)=0 and press the **+1 generation** button you get q'=0.1802.

Questions:
Calculate the gene frequency in the first 5 generations of selection and the equilibrium gene frequency in the genetic system shown below with selection favouring the heterozygotes with $s_1 = 0.3$ and $s_2 = 0.5$

```
Genotype          AA              Aa              aa        Total
Observed numbers 225             157              43
Frequency         p²              2pq             q²        = 1,00
Fitness           1-s₁            1               1-s₂
Proportion        p²(1-s₁)        2pq             q²(1-s₂)  = 1-p²s₁ - q²s₂
after selection
```

The frequency row uses p^2, $2pq$, q^2; fitness uses $1-s_1$, 1, $1-s_2$; proportion uses $p^2(1-s_1)$, $2pq$, $q^2(1-s_2)$ = $1-p^2 s_1 - q^2 s_2$.

How many generations will it take to reach equilibrium within 0.02 units ?

What would happen after 5 generations selection if $s_1 = -0.3$ and $s_2 = -0.5$
Would there be any equilibrium ?

3.5 Selection: Change in gene- and genotype frequencies, and effect of population size

In all cells with initial values you can put data (green and/or yellow fields)
You can put either real observations (numbers) or relative genotype frequencies in the green fields. If you see this text you can't run this applets

Click the **Calculate** button. Put the fitness of the genotypes (relative or numbers) and you can run one generation at the time by pressing the button **+1 generation!**, or you can get a graph over gene frequency changes over 60 generations by pressing the **Run** button. The corresponding data you can extract over the clip board in the lower left window.
If you define population size (pop. size) you can simulate random changes in gene frequencies over time. If you define more populations (number of pop.) you get an average estimate of the gene frequencies over generations and its standard error.
If you use large and many populations the simulation might take several minutes, start with few and small populations so you can have an impression of the time consumption on your computer.

For utilization of the deterministic model the upper part of this applet go to applet 3.4 and its explanation.

Example:

If you have a genetic system (Aa) with gene frequency q=.5. and you have random mating. How many populations out of ten would either have lost or fixed the gene a after 20 generations ? after 40 ? and after 60 generations ?

Put 10 in the pop size box, and pres **Run** 10 times recording loss or fixation of gene 'a' in each run.

Questions:
Calculate the gene frequency in the first 60 generations of selection in the genetic system shown below with selection favouring the heterozygotes with $s_1 = 0.3$ and $s_2 = 0.5$ having a population size of 20.

```
Genotype          AA              Aa              aa          Total
Observed numbers  225             157             43
Frequency         p²              2pq             q²          = 1,00
Fitness           1-s₁            1               1-s₂
Proportion        p²(1-s₁)        2pq             q²(1-s₂)    = 1-p²s₁ - q²s₂
after selection
```

Will equilibrium within 0.02 units be reached ?

4.4 An Applet for general matrix handling and calculating relationship and inbreeding

The relationship and inbreeding, and the inversion part of the matrix method has been programmed by Nils Toft and Thomas Nejsum Madsen, respectively, whereas the general matrix part has been borrowed from Bryan Lewis's General Matrix Applet. Link other applications of the General Matrix Applet: Estimating animal model breeding values and calculating selection index. If you see this you can't run the applet.

Matrices are just defined by line length and number of lines. To edit data in the data field is very troublesome as the return key can not be used for giving a new line as it works as pressing the eval. button. For larger sets of data use notepad or any other editor. Move the data to the input area by means of the clip board using ctrl-ins (copy) and shift-ins (insert).

Example:
Example of calculating inbreeding where the columns in the matrix shown below are animal, sire and dam, where the oldest animals should appear first. The pedigree are converted in the 'a' matrix as done below (0 means unknown parent):


```
animal sire dam
---------------
 a =
 1      0    0
 2      1    0
 3      1    0
 4      1    0
 5      2    3
 6      4    3
 7      5    6
```
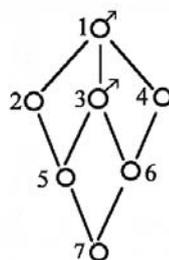
Take these data including (a=) in the clip board and transfer them to input window and there after press the 'evaluate' button and the matrix should appear in the result window.
There after use in the input window:

```
    rel = :tabular(a)
```
and press the 'evaluate' button

The rel relationship matrix will appear in the result window and will look as shown below. Number of digits can be set, replacing the 6 with 3 will change the number of digits to 3.

```
Relationship matrix  1+F (inbreeding) in the diagonal
 rel =
      1.000     0.500     0.500     0.500     0.500     0.500     0.500
      0.500     1.000     0.250     0.250     0.625     0.250     0.437
      0.500     0.250     1.000     0.250     0.625     0.625     0.625
      0.500     0.250     0.250     1.000     0.250     0.625     0.437
      0.500     0.625     0.625     0.250     1.125     0.437     0.781
      0.500     0.250     0.625     0.625     0.437     1.125     0.781
      0.500     0.437     0.625     0.437     0.781     0.781     1.218
```

The inbreeding for each animal is given by its diagonal element minus one, i.e. the last three animals are inbred with an inbreeding coefficient of 0.125, 0.125 and 0.218 respectively.
The identifying numbers of each animal can be any number, as long as the oldest animals appear first in the list. When defining the base (oldest) generation both parents are set as zeros (0 = unknown).

If you are interested in the inbreeding coefficient, only, you run the inbred procedure as follow:

```
     inb = :inbred(a)
```
you will have the following, in the results window, with animal number and inbreeding coefficient:
```
     Animal    F (inbreeding)
     inb =
      1.000     0.000
      2.000     0.000
      3.000     0.000
      4.000     0.000
      5.000     0.125
      6.000     0.125
      7.000     0.218
```

For inverting the relationship matrix use the command:

```
     inv = :invert(rel)
```

Or setting up the inverse relationship matrix direct from the a matrix only to use with non inbreeding:

```
     d_inv = :dinvers(a)
```
Which put the inverse relationship matrix in the output window

Below is shown a hand calculated example with the two general formula in use.
The relationship between two animals is the average relationship between the oldest and the others two parents.
and the inbreeding of an individual is half the relationship between the two parents.

$$a_{51} = \tfrac{1}{2}[a_{21} + a_{31}]$$

$$\tfrac{1}{2} = \tfrac{1}{2}[\tfrac{1}{2} + \tfrac{1}{2}]$$

Sire Dam

Parents	–	1–	1–	1–	3-2	3-4	5-6
Animal	1	2	3	4	5	6	7
1	1	1/2	1/2	1/2	1/2	1/2	1/2
2	1/2	1	1/4	1/4	5/8	1/4	7/16
3	1/2	1/4	1	1/4	5/8	5/8	5/8
4	1/2	1/4	1/4	1	1/4	5/8	7/16
5	1/2	5/8	5/8	1/4	1+1/8	7/16	25/32
6	1/2	1/4	5/8	5/8	7/16	1+1/8	25/32
7	1/2	7/16	5/8	7/16	25/32	25/32	1+7/32

$$R_{56} = \frac{a_{56}}{\sqrt{(1+F_s)(1+F_6)}} = \frac{7/16}{1+1/8} = 7/18 \qquad F_6 = \tfrac{1}{2}a_{34}$$
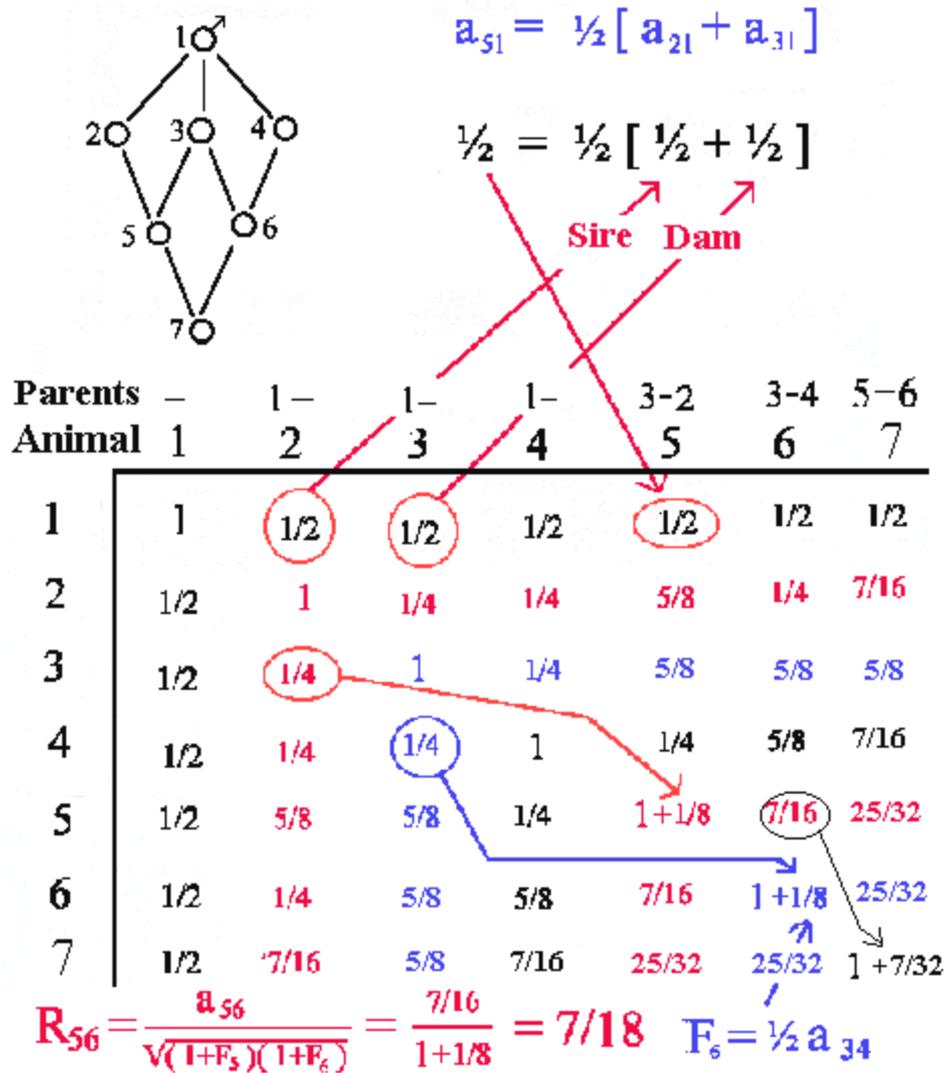
For your own data put them in the matrix 'a' and repeat the process:

Questions:
In the pedigree shown below the individuals have been assigned the numbers from 1 to 11.
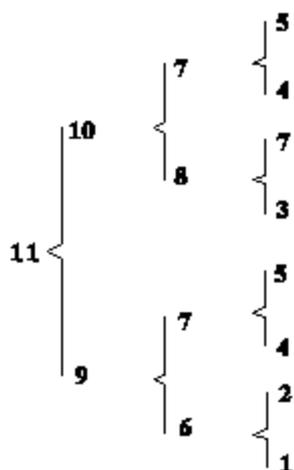

```
Animal Sire Dam   , where 0 means
unknown
-------------------
   a=
   1       0    0
   2       0    0
   3       0    0
   4       0    0
   5       0    0
   6       2    1
   7       5    4
   8       7    3
   9       7    6
  10       7    8
  11      10    9
```

Calculate the inbreeding coefficient for the individuals No 10 and 11, and calculate the additive relationship coefficient between individual 11 and all others in the pedigree. The data set to be used for the calculations is given above.

What would be the inbreeding coefficient after 5 generation of full sib mating?

5.6 Calculation of Chi-square test for deviation from Mendelian ratios

- Put your observed numbers and expected ratios (shall sum up to one) and click the **calculation** button. In all cells with initial values you can put data (green fields). if you see this you can't run the applet.
The number of degrees of freedom for the Chi-square is equal to the number of cells with expected values minus one.

Example:
Test of segregation ratios from known mating can be done by Chi-square test. If you have a test mating between two heterozygotes Aa x Aa the offspring would have an expected segregation ratio of 1:2:1 as shown in the table below, where the ratios are converted to a 0.25 : 0.5 : 0.25.

```
Genotype        AA              Aa              aa        Total
-------------------------------------------------------------
Number, obs     30              51              39      = 120 = N
Frequency,      .25             .5              .25     = 1,00
Number,         30              60              30       = 120
Deviations      0               -9              9
Chi-Square      0               1,35            2,70    = 4,05
----------
```
To test the applet put the bold face values from the table in the applet and press the **Calculate** button.

The number of degrees of freedom is df = 3-1 = 2, as the material is only providing the parameter N being used to calculate the expected numbers. The Chi-square value is less than the test value 5,99, which means that there is not statistically significant deviations from a 1:2:1 segregation ratio on the 5% level.

Questions:
Calculate a Chi-square for the following observation set from test mating of known carriers.

```
Total numbers of observations   30
Number of affected              16
```

Is the segregation ratio statistically significant different from a 1:3 segregation ratio ?

5.7 Calculation of corrected segregation ratio according to Singles method

The method is to be used when your data are collected in the field, and you only find useful families where at least one affected offspring occur. if you see this text you can't run this applets

In the cells with green colour put your data and press the **Calculate** button.

Example:
The disease tyrosinemia in mink is investigated by the following set of field data for recessive Mendelian inheritance. A description of the disease tyrosinemia in mink are published by Christensen et al. Canadian J. Comparative Medicine, 43:333-340, 1979.

In the data set 16 litters are included with at least one affected offspring. The observed numbers are given below.

```
T = 94    - total numbers of pups
A = 32    - total number of affected
A₁ = 4    - number of families with one affected pup
A₂ = 9    - number of families with two affected pups
```

The applied formulas for test of Mendelian inheritance by "Singles method" (with the proportion p-hat) are shown in the figure below. The tyrosinemia data are applied to show the use of the formulas.

$$\hat{p} = \frac{A - A_1}{T - A_1} = \frac{32-4}{94-4} = 0,311$$

$$Var(\hat{p}) = \frac{T - A}{(T - A_1)^3}\left[A - A_1 + 2A_2\frac{T - A}{T - A_1}\right]$$

$$= \frac{94-32}{(94-4)^3}\left[32-4 + 2*9\frac{94-32}{94-4}\right]$$

$$= 0,0034$$

$$Z^2 = \frac{(\hat{p} - p)^2}{Var(\hat{p})} = \frac{(0,311-0,25)^2}{0,0034} = 1,10$$

The Z^2 are Chi-squared distributed with 1 degree of freedom. The test of p-hat is against 0,25, this is the expected ratio when both parents are carrier and there is recessive inheritance. The test value is less than 3.84. Therefore the deviations can be accepted as not statistically significant.

Questions:
Calculate the corrected segregation ratio for the following observation set of field data, where there has been at least one affected per family.

```
T =  30   - total numbers of observations
A =  16   - total number of affected
A₁ = 12   - number of families with one affected
A₂ =  2   - number of families with two affected
```

Is the segregation ratio statistically significant different from 1:3 segregation.

Applet for mean, Genotypic and Breeding values and Dominance

6.2 Applet for calculation of mean, Genotypic and Breeding values and Dominance deviations

In all cells with initial values you can put data (green and/or yellow fields)

if you see this you can't run the applet.

When you have entered your data or you want to calculate the loaded test example press the **Calculate** button, you can proceed by setting deltaq, or getting the graph by pressing the run button and you get the new values by pressing the **q+deltaq** button or getting the graph by pressing the run button.

Example:
The test example is the transfferin example given in Veterinary Genetics. Jersey cows with the genotype tt yield 2082 kg milk and cows with the genotypes Tt and TT yield 1882 kg milk. The gene frequencies p(T) and q(t) are given as 0,67 and 0,33, respectively.

All the values are calculated by the classical formula for a mean value and the breeding value is defined as shown below. To see the results press the **Calculate** button.

$$A = 2(\bar{P}_g - \bar{P})$$
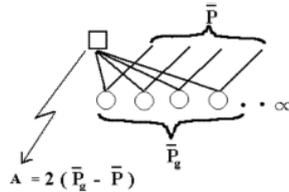
Questions:
Calculate the breeding value and dominance deviation for the gene frequencies p equal .1, .2, .3, .4, .5, .6, .7, .8 and .9 for the following genotypic values

```
A₂A₂    100
A₁A₂    115
A₁A₁    110
```

Compare the results, for which gene frequency are there the smallest difference in breeding value between the three genotypes and the smallest additive variance? why ?

The halothane locus (Hal) has influence on the quality of swine meat on the basis of a meat quality index where meat colour is the main component. The average index for the (Hal) genotypes are given below:

Genotype	Meat-quality-index
N/N	8.00
N/n	7.10
n/n	6.00

Suppose that the frequency of $q(n) = 0.20$ calculate the following:
a) the average Meat-quality-index for the population,
b) the average Meat-quality-index for offspring from individuals with the genotype N/N respectively n/n supposing random mating in the population,
c) the breeding value of the three genotypes
d) the additive variance, due to the halothane locus,
e) the proportion of the additive variance$_{Hal}$ of the total additive variance for the Meat-quality-index as its $h^2 = 0,44$ and variance$_P = 1,69$.

7.2 Estimating simple forms of breeding values

The breeding value can be based on own, fathers + mothers record as well as average of offspring or sibs to the animals. Heritability, repeatability, common environmental effect (milieu) for full sibs and population mean value have to be specified. In the last case relationship (a and a') shall be defined. The relationship (a) between the related individuals giving the phenotypic measurements; with the relationship (a') to the animal being evaluated. The 'common milieu' is used when full sibs are measured.

If you see this you can't run the applet.

In all cells with initial values you can put data (green fields) and press the **Calculate** button. Results are rounded to four decimals.

Example:
If you want to see how the applet work before you enter your own data just press the **Calculate** button using the initial values.
In the figure below you see a regression line for the index (I) for the true breeding value (A) of an individual based on measurements (average of the records for offspring, sibs etc. = P's).. The coefficients of relationship can be defined in the bottom line.
P'er = phenotype measurements

112

n = number of phenotypes
P_g bar = average of the P's
a' = relationship between the index animal and the P's
a = relationship between the P's
P bar = Population mean
A bar = Population mean A's = P bar
h^2 = heritability
c^2 = common environmental factor for the P's
r_{AI} = the accuracy of the index
$t = a\,h^2 + c^2$ = phenotypic correlation between P's

$$I = \bar{A} + \frac{n \cdot h^2 \cdot a'}{1+(n-1)\,t}(\bar{P}_g - \bar{P})$$

$$\bar{P} = \bar{A}$$

$$\frac{\Sigma P}{n} = \bar{P}_g$$

$$(\bar{P}, \bar{A})$$

$$\sigma^2_{A.I} = \sigma^2_A(1 - r^2_{AI})$$

$$r^2_{AI} = \frac{n \cdot h^2 \cdot a'}{1+(n-1)\,t}\,a'$$
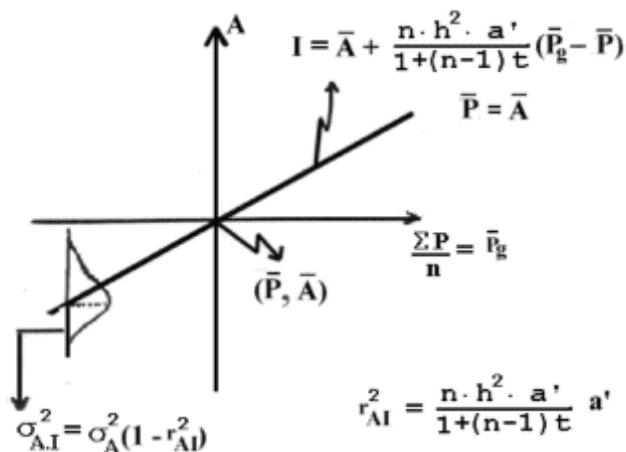
Questions:
Calculate the breeding value for the following animals when Mean=100 and the heritability have the following values 5, 10, 25, 40 and 60 %.

```
own performance        110
average of the parents 115
10 half sib offspring  106
100 half sibs          105
```

Compare the results, which animal have the highest breeding value for each heritability? And which one have the most accurate estimate for each heritability ?

7.2 Estimating animal model breeding values

The calculations are based on the general matrix part which has been borrowed from Bryan Lewis's General Matrix Applet. The general part include for instance multplication as A=b*c or A=b*c' which can be used together with the the animal model calculations.

The applet can handle inbreeding, relationship, calculating the inverse and **estimating animal model breeding values.** It will perform well up to 2000 observations and might be even more. It has been made for solving smaller scientific problems and for educational purposes. It is worth to note that the numeration of the animals is free as long as the oldest appear first. Link example code=TabularApp.class name=TabularApp width=800 height=500 >If you see this you can't run the applet.

The estimating the animal model breeding values, command an=:animalm(a), can correct for any number of class variables or continued variables in any combinations.
For class variables the fist class is excluded for not having a degree of freedom as the mean value always is calculated. Zero is treated as missing value so if zero is a part of the observation set a constant should be

added. The first 3 columns in the dataset shall contain animal, sire and dam - all other columns can be used freely. [Link example](#)

The first data line specify the parameters as follows - referring to first example below as well as in general

```
1.  1 specify (sigmaE/sigmaA)²  which in this case equals h² = .50

2.  0 printing the solutions
    1 printing the solutions and the accuracy squared, for smaller number of data (pc-
time problems)
    2 printing the triangualrization inclusive dependent var - last line the solutions
    3 printing the rearranged input with classes from zero to n-1 and dependent
variable as the final.
    4 printing the equations inclusive the dependent var.

3.  0 normal model
    1 model with inbreeding - for smaller number of data (more roundings and pc-time
problems)

4.- 0 exclude the trait from the analyses
    1 specify a continued variable where the last one is the dependent trait
   -1 specify the dependent trait if not the last trait -  which is then interchanged
with the last variable
    2 specify a class variable
    3 specify a class variable but with all classes included (this can not be solved)
```

Example of estimating animal model breeding values for 8 animals, where the input file contains: Animal Sire Dam, where 0 means unknown, one fixed factor, two continued vars. and the trait litter size - except first line which specify the parameters explained above.

Put the data below in **the input window** followed by a return stroke

```
a=
1 0 0 2 1 1  1
1 0 0 3 2 0 10
2 0 0 2 3 1  9
3 0 0 1 4 0  8
4 0 0 2 1 1  7
5 1 2 1 5 1  9
6 1 2 1 6 0 10
7 3 4 2 4 1  8
8 5 6 3 6 1 11
```

Genealogical diagram
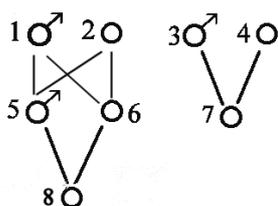
Resulting in the following in the **output window**

```
a =
1  0  0 2 1 1  1
1  0  0 3 2 0 10
2  0  0 2 3 1  9
3  0  0 1 4 0  8
4  0  0 2 1 1  7
5  1  2 1 5 1  9
6  1  2 1 6 0 10
7  3  4 2 4 1  8
8  5  6 3 6 1 11
```

Followev by   **an=:animalm(a)** and a return stroke resulting in

```
Animal model solutions
an =
 1.0000  0.1955
 2.0000  0.4287
 3.0000 -0.3799
 4.0000 -0.2443
 5.0000  0.3420
 6.0000  0.3582
 7.0000 -0.4206
 8.0000  0.2903
 0.0000  6.9749 mean value
 2.0000  0.5183 class  there is
no degree of freedom for class
1
 3.0000  1.8716 class
 0.0000  0.4191 regression 1
 0.0000 -0.5322 regression 2
```

In the referred theory by Poul Jensen the first example on growth rate in young pigs can not be solved as the class year is given for all animals as follows. But the system will nullify the first year and it is then converted to a regression giving 4.3883 gram extra the second year. The heritability is set to 0.33.

```
a=                Genealogical diagram                       Animal model solutions
```

```
2 0 0 2 1
1 0 0 0 0
2 0 0 1 225
3 0 0 1 220
4 0 0 1 255
5 1 3 2 250
6 1 3 2 198
7 2 4 2 245
8 2 4 2 260
9 2 4 2 235
```
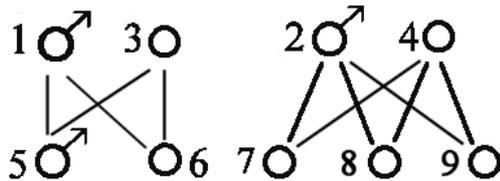


```
animal    estimate
an =
 1.0000 -3.1844
 2.0000 -0.3009
 3.0000 -6.2135
 4.0000  9.6990
 5.0000 -1.0912
 6.0000 -11.4912
 7.0000  5.4271
 8.0000  8.4271
 9.0000  3.4271
 0.0000  232.2718
 0.0000  4.3883
```

In case some of the parents are inbred the estimates will not be the same. In the following small example solutions are made with (parameter 3 set to 1) and without inbreeding. Example 2.1 from RA Mrode 'linear models for estimating breeding values, CABI Publishing'. The heritability is set to 0.33.

```
a=
2 0 1 1
1 0 0 15    Genealogical
2 0 0 19    diagram
3 1 2 7
4 1 0 13
5 4 3 22
6 5 2 10
```


Genealogical diagram
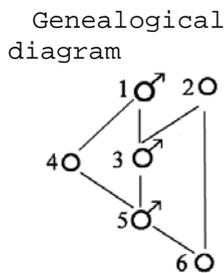
```
Animal model solutions

not accounting for
inbreeding
an =
 1.0000 -0.3699
 2.0000  0.1725
 3.0000 -1.0988
 4.0000  0.1110
 5.0000  0.7084
 6.0000 -0.5484
 0.0000  14.5041
```

```
Animal model solutions

accounting for
inbreeding
an =
 1.0000 -0.3665
 2.0000  0.1689
 3.0000 -1.0983
 4.0000  0.1131
 5.0000  0.7070
 6.0000 -0.4987
 0.0000  14.4957
```

Estimating of the accuracy squared for the estimate; example 3.1 RA Mrode. The heritability is set to 0.33.

```
a=
2 1 0 2 1
1 0 0 0 0
2 0 0 0 0
3 0 0 0 0
4 1 0 1 4.5
5 3 2 2 2.9
6 1 2 2 3.9
7 4 5 1 3.5
8 3 6 1 5.0
```

Genealogical diagram
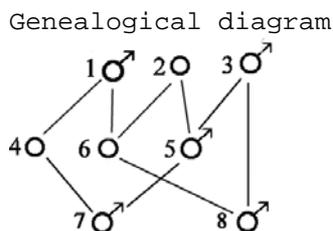

```
Animal model solutions
with its squared accuracy
an =
 1.0000  0.0984  0.0578
 2.0000 -0.0187  0.0158
 3.0000 -0.0410  0.0870
 4.0000 -0.0086  0.1446
 5.0000 -0.1857  0.1437
 6.0000  0.1768  0.1154
 7.0000 -0.2494  0.1162
 8.0000  0.1826  0.1552
 0.0000  4.3585  0.0000
 0.0000 -0.9540  0.0000
```

Larger example solutions for y, **note that there is free numbering of the animals** as long as the oldest occur first, sex 2 has a 5.6002 units lower performance.

Animal model solutions

The first variable contains numbers

```
animal sire  dam   y  sex trait2        of animal (mean) classes and regressions
a=                                      and the second the estimate
2 0 0 -1 1 2                            an =
1 0 0 0 0 0                              1.0000 -1.4325
2 0 0 0 0 0                              2.0000 -4.4927
3 0 0 0 0 0                              3.0000  3.1002
4 0 0 0 0 0                              4.0000  0.4142
5 0 0 0 0 0                              5.0000  2.9162
6 0 0 0 0 0                              6.0000 -3.3706
7  0  0   93   2 2                       7.0000  2.8651
8  1  2   78   1 2                       8.0000 -4.7612
9  1  2   63   1 3                       9.0000 -5.6569
10 1  3   91   1 3                       10.0000      2.9802
11 1  3   70   2 3                       11.0000     -0.0996
12 1  3   100   2 2                      12.0000      2.7212
13 1  4   102   1 4                      13.0000     -0.0949
14 1  5   119   1 5                      14.0000     -0.4850
15 1  5   121   2 4                      15.0000      4.8850
16 1  6   93   2 6                       16.0000     -3.7721
17 1  6   81   2 4                       17.0000     -6.5152
18 1  6   82   2 3                       18.0000     -0.2880
19 14 17 78   2 2                        19.0000     -4.0711
20 14 17 88   1 5                        20.0000     -5.4854
21 14 17 85   2 6                        21.0000     -5.5097
22 14 7  91   2 5                        22.0000     -0.0535
23 14 7  113   1 7                       23.0000     -1.6506
24 14 7  118   2 5                       24.0000      5.2497
25 14 15 91   2 4                        25.0000      0.6864
26 14 15 83   2 2                        26.0000      1.4889
27 14 15 102   1 4                       27.0000      2.0724
128 14 12 87   2 5                       128.0000     -3.1481
29 14 12 83   1 3                        29.0000      1.6076
30 14 12 86   1 3                        30.0000      2.2076
31 14 16 106   2 4                       31.0000      0.8187
32 14 16 90   1 5                        32.0000     -3.9882
33 14 16 57   2 3                        33.0000     -5.0696
34 23 25 82   1 2                        34.0000     -1.9767
35 23 25 84   1 1                        35.0000     -0.4821
36 23 25 77   2 3                        36.0000      0.2475
37 23 22 83   2 2                        37.0000     -0.9527
38 23 128 95   1 5                       38.0000     -3.2048
39 23 128 68   2 4                       39.0000     -7.2870
40 23 31 116   1 6                       40.0000      2.0375
41 23 31 112   1 7                       41.0000     -0.6642
42 23 31 88   2 5                        42.0000     -1.8980
43 23 24 121   1 4                       43.0000      5.5520
44 23 24 98   1 4                        44.0000      0.9520
45 23 24 86   1 4                        45.0000     -1.4479
                                         0.0000  90.0823
                                         2.0000   5.4732
                                         3.0000  -5.0481
                                         4.0000  15.9558
                                         5.0000  16.9446
                                         6.0000  19.6661
                                         7.0000  29.1753
                                         0.0000  -5.6002
```

The theory has been given by late Poul Jensen; See also [Erling Strandbergs notes](#)

8.2 Estimating breeding values and selection response

The breeding value can be based on any uniformly related individuals to the animals or own record, see applet 7.2, remember the actual relationships a and a' should be properly defined. By pressing the Calculate button the accuracy is transferred to the delta G input window. if you see this you can't run the applet.

In all green fields you should define values then press the **Calculate** button. The delta G can be run independently of the breeding value estimations, all values in the yellow fields shall be defined.

The selection response is calculated in the lower 3 rows of the applet according to the formula shown below. The delta G is calculated per year (år). If the generation interval is set to one year the results is also per generation. The parameters are
Selection intensity, i
Accuracy of selection, r_{AI}
Generation interval, L
The additive standard deviation, $sigma_A$

$$\Delta G_{year} = \frac{i_{\female} r_{AI_{\female}} \sigma_{A_{\female}} + i_{\male} r_{AI_{\male}} \sigma_{A_{\male}}}{L_{\female} + L_{\male}}$$

Example:
The accuracy is automatically transferred from the breeding value estimation, but you can also convert an externally derived accuracy squared by means of r^2 to r button. If you have the proportion selected (p) put your data in the (i) field and press the convert p to i button.
The calculation is based on the average of the contribution from the sires and dams. With the given initial parameters delta G is 0.04 units if you press the **Calculate delta G** button.

Questions:
In a population are selected 10 % males based on progeny test of 20 offspring and 50 % females based on own performance. SigmaA is set to 1.
Calculate delta G per generation when the heritability is 0.40.

If the males as well as the females were selected on own performance calculate delta G.
Compare the two results

Repeat the calculations with a heritability of 0.10

Applet for calculating heritability for threshold traits.

8.3 Applet for calculating heritability for threshold traits (diseases)

The calculations are based on disease frequencies in the general population (Population) and among relatives (Indiv. related) to diseased animals. The relatives can be (both parents diseased), first degree (father, mother or full sibs) or second degree (half sibs, grand parents).
Observe: The chi-square test calculation is based on: that the related ones is subtracted from the population numbers!

Example:
Enter your observed numbers and degree of relationship (green fields) and then press the **Calculate** button. The values in the (Degree of relationship) field should 1, .5 and .25 for the three above mentioned cases. You can also enter the frequencies (yellow fields) but then the numbers shall be zeros in the observed number fields. Results are rounded to four decimals.

Illustration of the case where both parents are affected are given below. The applet is initiated with the actual data; to see the calculations press the **Calculate** button.

Population with 5 % affected

Basic formula
$$R = h^2 S = h^2 i$$

5 %

$\overleftarrow{X_1}\rightarrow$
$\leftarrow i = S \rightarrow$

Offspring from affected parents have 20 % affected

$$h^2 = \frac{R}{i}$$
$$= \frac{X_1 - X_2}{i}$$
$$= \frac{1.6 - 0.8}{2.06} = 0.39$$

20 %

$\leftarrow R \rightarrow \overleftarrow{X_2} \rightarrow$
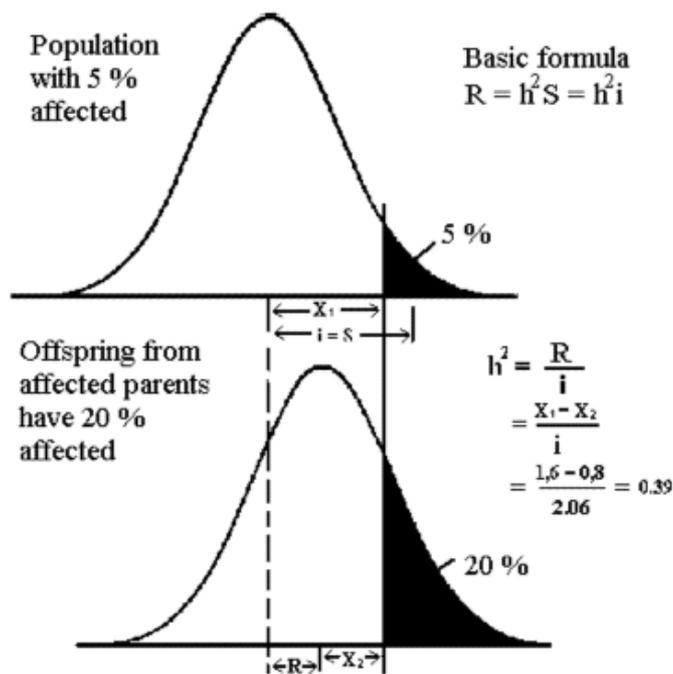
Remember in case of first or second degree relative analyses with half or full sib, one affected individual per family is an index case and should be subtracted from the group 'individual related'. So if there are 10 families with at least one affected in the population, 10 affected should be subtracted.

Questions:
In the table below is given observed numbers for disease frequencies in the population and in first-degree relatives to diseased animals.

	Normal	Diseased	Frequency
Population	980	20	0.020
First-degree relative	140	10	0.066

Estimate the heritability.
What is the relative risk to be related to a diseased animal ?
Is the heritability statistically significant different from 0 with ($p < 0.05$) ?