Selection and Genetic Change

2006-06-14

Erling Strandberg and Birgitta Malmfors

Dept of Animal Breeding and Genetics
Swedish University of Agricultural Sciences, Uppsala, Sweden

Contents

Selection – the major tool for genetic improvement 1
Natural and artificial selection 1
Selection gives long-lasting results 2
Selection strategies 2
Selection for several traits 4
How to predict genetic change from one generation to the next 5
Prediction of genetic change when selecting on individuals' phenotypes 6
Prediction of genetic change from the proportion selected 8
General equation for genetic change from truncation selection 9
How to predict genetic change per year 13
The components of genetic gain are inter-related 14
Prediction of correlated genetic change 16
Realized genetic change 17
Selection and genetic variation 18
General results from long term selection experiments 20
Concluding remarks 21
References 22
Some literature for further reading 22

Appendices 1-8 are in a separate document

Selection and Genetic Change. Compendium, version 2006-06-14

Selection – the major tool for genetic improvement

The ultimate goal of a breeding program is genetic improvement of traits defined in the breeding objective for the animal population. The major tool to achieve this is to select the best animals as parents to produce the next generation, and among those parents also decide which ones should have the largest number of offspring. With successful selection, the progeny generation will on average be better than the average of the population from which the parents were chosen – a genetic progress is obtained. Selection does not create new genes, but increases the proportion of favourable genes in the population. The gains are accumulated when we continue to select the best animals in each generation. A continuous, long-lasting genetic improvement in traits included in the breeding goal is thus achieved.

What genetic change can be attained depends on a number of factors. First of all, the traits that are to be improved must show additive genetic variation, and we need to be able to identify the best animals. It also matters how many traits we include in the breeding goal, what proportion of animals that are selected, how intensively they are used, how long the generation interval will be, et cetera. Moreover, we need to be aware of the potential existence of genotype-environment interactions, and that a specific genotype is not always the best one in all environments. A good knowledge and understanding of various factors influencing genetic progress and how they can be optimised is crucial for us to be able to fully utilize our most important tool in animal breeding, selection!

Natural and artificial selection

Selection is a normal phenomenon occurring in all kinds of living materials - it is the mechanism behind evolution. In nature it is the animals that are best adapted to their environment that survive and produce the largest number of progenies, in other words, it is first of all the animals that are the "fittest" that contribute with genes in the population. This is what we call ***natural selection***, and it mainly favours a combination of viability and reproductive ability, i.e. fitness traits.

The selection done under human control to obtain genetic improvement of traits in domestic animals is called ***artificial selection***. Fitness traits, such as fertility and disease resistance are usually included in this selection, but large emphasis is given also to many other traits, such as production traits, productivity, product quality, performance traits and longevity. Natural selection occurs simultaneously with the artificial selection, but maybe not as forceful as in nature, because domestic animals are usually not exposed to the same harsh environment as are wild animals. Natural and artificial selection may support the same goal, but they might also act in opposite directions. This might be the case if fitness traits are not given enough emphasis in the artificial selection. We will from here onward focus on artificial selection only, which will simply be referred to as selection.

Selection can be performed both between and within populations (e.g. breeds). To screen animal populations and thereafter use those that have characteristics in line with a desired breeding goal can be a way to get results quickly, assuming the populations can be compared properly. For continuous and long-lasting effects, however, it is necessary to conduct selection within populations. This is what is normally meant by selection for genetic improvement.

Selection gives long-lasting results

Selection and efficient use of the selected animals, normally results in an increased proportion of desirable genes in the population, and a subsequent genetic improvement of the population means for the traits under selection. The genes from the selected animals will be spread also to future generations, so the genetic gain obtained through selection will last. The latter is true as long as natural selection or any correlated, undesirable, changes will not counteract the gain achieved. Recurrent selection over several generations results in genetic gains that are accumulated, as every round of selection starts from the level that was obtained in the previous round. The effect of selection is illustrated in Figure 1.

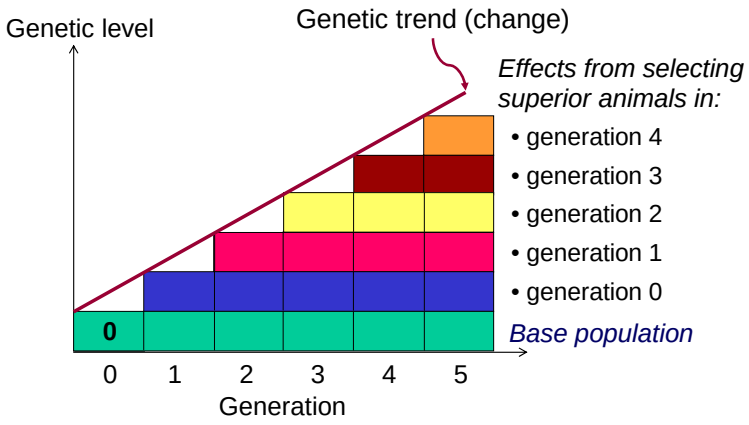


Figure 1. Illustration of how selection effects are accumulated and maintained.

From the figure we can see that:

- Using selected animals from the base population as parents results in an increased genetic level of the animals in the next generation (see the column for generation 1).
- The selection effect obtained through selection in generation 0 is maintained also in later generations (see the row for selection effects from generation 0).
- Selecting superior parents also in subsequent generations further raises the genetic level of each generation.
- The selection effects from each generation are accumulated. Thus, the genetic level of animals in generation 5 is built up by selection effects from all previous generations (see the column for generation 5).
- The genetic trend per generation is illustrated through the line.

The driving force for genetic improvement is of course the genetic superiority achieved by the selection of parents, but also other factors, such as generation intervals and differentiated use of selected breeding animals, have an impact. One should also remember that additive genetic variation is a prerequisite for any selection effort to be successful. The animals in the population must be genetically different with regard to a specific trait. Otherwise it will not be possible to select individuals that are better than others!

Selection strategies

The traits we want to improve in a population are defined in the breeding goal. A few of the goal traits might be influenced by simply one or a few alleles, which means that the true genotype often can be determined (e.g. through a DNA test) and it is then easy to select the desired genotypes. The majority of goal traits, however, are quantitative in nature, i.e. influenced both by genes at many loci and

by environment. These traits are often normally distributed. Selection is then commonly based on predicted breeding values, which in turn are based on phenotypic values and knowledge of heritabilities, genetic correlations, genetic relationships, economic weights, etc. By using the BLUP Animal Model for the genetic evaluation we can utilize all phenotypic information available, and we get breeding value predictions that are also properly adjusted for systematic non-genetic effects (see the compendium "Genetic Evaluation".

Selection on quantitative traits can be either ***directional*** or ***stabilizing***. The two types are illustrated in Figure 2.

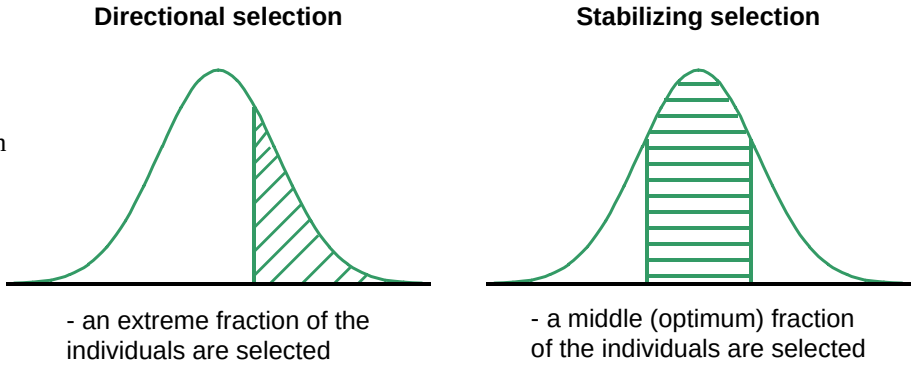


Figure 2. Illustration of directional and stabilizing selection.

Directional selection is the most common type of selection. It means that an extreme fraction of the individuals are selected. If a high value for a trait is desirable, then we select the animals with the highest values, e.g. those with high growth rate, milk yield or performance score. If a low value is desirable, then we select animals in the opposite fraction of the normal distribution, i.e. the animals having low values, e.g. for back-fat thickness, disease incidence or time to run a race.

Stabilizing selection means that we select a middle fraction of the animals and avoid selecting the extremes. In this type of selection it is the optimum values that are desirable. Examples could be birth weight and quality traits, such as meat tenderness or curl-size of pelt skins. It is possible also that in species with large variation in litter size one wants to avoid selecting animals giving very small or very big litters. From here on we will concentrate on directional selection.

Selection within a population is usually applied in several stages; this is sometimes called ***stepwise selection***. The first selection event might be based entirely on pedigree information (usually the average of the parents' breeding values). The next events might occur when information is available on animals themselves, and maybe also on sibs; one selection round on traits expressed fairly early in the animals' lives, e.g. growth rate, and another round on traits expressed later, e.g. fertility or performance. A fourth selection round might occur when in addition to previous information there are also progeny results at hand. There might also be one step for selection of elite animals as parents to the next generation of males to be used in artificial insemination, for example. The best animals in each round of selection are retained to the next selection event, while the ones that are not selected might be culled, or they might be used as production animals, or even as parents to non-elite breeding animals.

Measuring and keeping records of important traits, and also predicting breeding values without bias and with a high precision, is fundamental for a functioning selection program. The accuracy by which we are able to rank the individuals determines the success of a selection among them (methods for prediction of

breeding values and their precision are described in the compendium "Genetic Evaluation"). As you will se later, the selection response (genetic change) is directly proportional to the accuracy of the genetic evaluation performed. Note that the accuracy may increase with time if more information is then available. With more information also the predicted breeding values of individual animals may change, i.e. either increase or decrease, which might lead to a change in ranking.

Selection for several traits

Selection programs in practice normally aim at improving several traits simultaneously. We want to select animals that have not only good production or performance, but also have good health and reproduction, etc. The situation is usually handled by predicting breeding values in a multiple-trait BLUP analysis or by a multi-trait selection index procedure. The BLUP breeding values (I_j) are thereafter weighted together by their relative economic weights (v_j) into a ***Total Merit Index (TMI)***, i.e. TMI = $v_1I_1 + v_2I_2 + \ldots.+ v_mI_m$.

Weighting traits together implies that a negative (unfavourable) value in one trait can be compensated by a positive value in another trait. For most cases this idea of weighting traits together is reasonable, but sometimes it is not acceptable. For example, a boar that has problems to stand on his feet should not be selected regardless of how good his TMI might be! An option can then be to combine the index selection with a method called ***independent culling*** (or threshold selection). In this method the animals must exceed a fixed minimum value for the trait (a threshold) to be acceptable selection candidates. Independent culling can be applied also for simultaneous selection of two or more traits. The two selection strategies are illustrated in Figure 3.

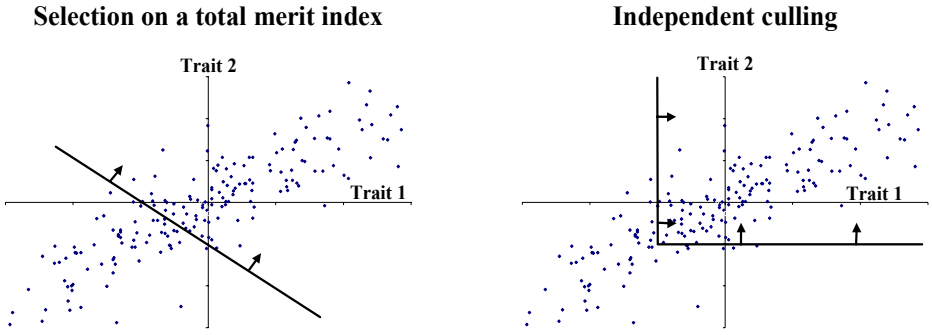


Figure 3. Illustration of two selection strategies: selection on total merit index and independent culling

We can see that it is not exactly the same animals that are selected with the two methods. For example, a few animals that are above the average for trait 2 but below the average for trait 1 will be selected if TMI is used, but they will not be selected with independent culling. On the other hand, some of the animals that will not be selected in the TMI case (those just below the TMI line in the left bottom part of the diagram) will be selected if, instead, independent culling is used. We could also consider selection for more than two traits. Any of the two methods could still be used, but the more traits that are to be considered, the more difficult it will be to use the method of independent culling. But the situation could also be that independent culling is used to guarantee a minimum value for one trait, and that the animals passing that threshold are thereafter ranked on a total merit index including the other two traits.

In the following we will discuss the main factors influencing genetic change in a population, and also show how the expected genetic change can be predicted.

How to predict genetic change from one generation to the next

Let us start by assuming that we know the true breeding values of all individuals. Then we can write the breeding value of an individual *i* as:

$$A_i = \frac{1}{2}A_s + \frac{1}{2}A_d + A_{MSsi} + A_{MSdi} \quad [1]$$

where A_s and A_d are the breeding values for the sire and dam, respectively, and A_{MSsi} and A_{MSdi} are the Mendelian sampling terms. In the parental generation we assume that the average breeding values are zero, for both males and females.

The breeding value of, say, the sire is based on what he is expected to inherit to the offspring, on average. However, any given offspring *i* will deviate more or less from this expectation because of the random sampling of genes from parent to offspring. This means that the offspring will actually get $0.5A_s + A_{MSsi}$, where you can think of A_{MSsi} as a (positive or negative) deviation from the expected value. This Mendelian sampling term will be different for each offspring. For any given individual we do not know the Mendelian sampling terms, we only know that in the long run they will be zero (on average). Usually, one only writes one Mendelian sampling term, but we just wanted to point out that there is one term from the father and one from the mother.

Each of the four components in [1] have an expectation of zero and make up one fourth of the total additive genetic variance (in the simplest case when there is no selection or inbreeding: ~IND$(0, 0.25\sigma_A^2)$. This also tells us that even if we know perfectly the breeding values of both sire and dam of an offspring group, we will only explain half of the genetic variation in the offspring – or expressed the other way around, half of the genetic variation is visible among the offspring of one parent couple. (This is why pedigree information cannot become as precise as progeny information, in the former case you can get a maximum reliability (r_{TI}^2) of 50% whereas progeny information can basically give 100% reliability.)

If instead of looking at one individual, we look at a group of offspring from two individuals, the average breeding value of that group will be:

$$\bar{A} = \frac{1}{2}A_s + \frac{1}{2}A_d \quad [2]$$

because the Mendelian sampling terms become zero on average.

Now, in reality we cannot observe the true breeding values, so we have to make do with the predicted breeding values. If we assume that we have unbiased predicted breeding values, i.e. $E(A|\hat{A}) = \hat{A}$, then for an individual we can write that the predicted breeding value is:

$$\hat{A}_i = \frac{1}{2}\hat{A}_s + \frac{1}{2}\hat{A}_d \quad [3]$$

The Mendelian sampling terms are not included because we cannot predict them for a given individual. Now, equation [3] refers to *one* individual, but we can do

the same for a group of individuals, and the average PBV for the group is equal to the average PBVs of all parents:

$$\bar{\hat{A}} = \frac{1}{2}\bar{\hat{A}}_S + \frac{1}{2}\bar{\hat{A}}_D \quad [4]$$

where subscripts S and D refer to the whole group of male and female parents, respectively. In genetic evaluation we often talk about ***index*** (*I*) instead of predicted breeding value and if we put that terminology into equation [4], we get the average index for the offspring group as:

$$\bar{I} = \frac{1}{2}\bar{I}_S + \frac{1}{2}\bar{I}_D = \Delta T_g \quad [5]$$

This is then the definition of ***genetic change per generation,*** ΔT_g (because we assumed that the average breeding value in the parental generation was zero). So, in principle, *the calculation of expected genetic change from one generation to the next, just comes down to knowing the predicted breeding values of the parents of the next generation* (or more strictly: how these PBVs deviate from the average predicted breeding value for the whole parental generation).

We will see how we can use this principle for all situations, but we start with the same simple example as when we started to predict breeding values – when we have one phenotypic observation on each individual (each potential parent).

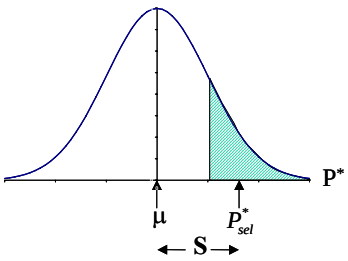
Prediction of genetic change when selecting on individuals' phenotypes

We have previously seen that the predicted breeding value of an individual *i* based on its phenotypic value is:

$$I_i = \hat{A}_i = b_{A/P}P_i = h^2 P_i = h^2(P_i^* - \mu) \quad [6]$$

where *P* is the phenotypic value as a deviation from the mean and *P** is the actual phenotypic value (possibly adjusted for some systematic environmental effects). Now, if we have a selected group of animals with a mean phenotypic value $\bar{P}^*_{sel}$ we can calculate the average predicted breeding value in the same way:

$$\bar{I}_{sel} = \bar{\hat{A}}_{sel} = h^2(\bar{P}^*_{sel} - \mu) = h^2 S \quad [7]$$



where we have defined a new term, ***S, the selection differential***, as the difference between the phenotypic average of the selected group ($\bar{P}^*_{sel}$) and the average of the group of animals it was selected from (μ). Consequently, the selection differential is a measure of *the phenotypic superiority of the selected group*.

Now, equation [7] describes how to calculate the average predicted breeding value of a selected group. Combining this with [5] we can get the genetic level of the offspring, which then is the genetic change per generation:

$$\Delta T_g = \frac{1}{2}\bar{I}_S + \frac{1}{2}\bar{I}_D \quad [8]$$

$$= \frac{1}{2}h^2 S_S + \frac{1}{2}h^2 S_D \quad [9]$$

where $\overline{I}_S$ and $\overline{I}_D$ are the average PBVs for selected males (sires) and females (dams), respectively, and similarly S_S and S_D are the corresponding selection differentials. Usually we have the possibility to select more intensively among males than among females, which makes $S_M > S_F$. Note, however, that males and females still contribute the same to the genetic setup of the offspring, half the genes come from the father and half from the mother.

Nevertheless, if we first for simplicity assume that both males and female parents are selected with equal intensity we get the genetic change from parent to progeny generation as:

$$\Delta T_g = \overline{I}_{sel} = h^2 S \qquad [10]$$

This genetic change from parent to offspring generation is illustrated in Figure 4. What we can see is that the genetic change from one generation to the next depends on the phenotypic superiority (S) of the animals that are selected as parents and on how much of this superiority that can be expected to be due to additive genetic origin, and thus transmitted to the offspring. The latter is indicated by the heritability (h^2). Another way to express the same thing is to say that the genetic change depends on the expected additive genetic superiority of the selected animals ($\overline{I}_{sel}$ which here is equal to $h^2 S$).

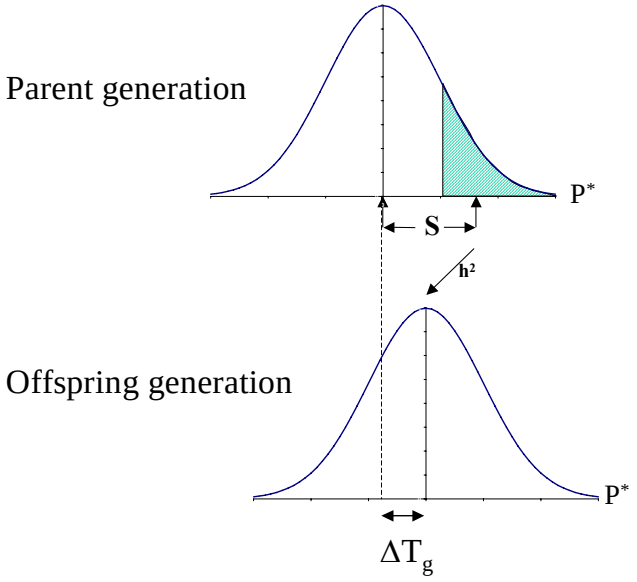


Figure 4. Schematic description of genetic change expressed as $\Delta T_g = h^2 S$.

In the slightly more general form with $b_{A/P}$ instead of h^2 the expression for ΔT_g applies also to other phenotypic measurements than the candidates' own performance. The regression coefficient to be used is the same as the selection index weight when predicting a candidate's breeding value using the phenotypic measure. A limitation with this approach to predict genetic gain is, however, that all candidates are assumed to have the same type and amount of phenotypic information. Therefore the version of the genetic gain equation including the selection differential applies only to rather simple situations whereas the more general approach,

$$\Delta T_g = \overline{I}_{sel} \qquad [11]$$

is always true.

Prediction of genetic change from the proportion selected

To be able to use the equations $\Delta T_g = b_{A/P} S$ and $\Delta T_g = h^2 S$ to predict the genetic change we need to know which animals are selected, and also their phenotypic values, so that we can calculate the selection differential (*S*). When planning a breeding program it would, however, be useful to be able to predict the genetic change for a certain selection strategy ***before*** we know which individuals that will be selected. Fortunately this is possible to do, when we practice truncation selection (i.e. when we select individuals above or below a certain truncation point, or threshold). Then the selection differentials can be derived theoretically from the distribution of the selection criterion (whether it consists of predicted breeding values or only phenotypic values) and knowledge of the proportion of selected individuals, which we will show below.

Let us again assume that we have the simple situation with one phenotypic measurement on each individual. We define the *standardized selection differential*, usually called the ***selection intensity***, as:

$$i = S / \sigma_P \qquad [12]$$

where *S* is in the units of the selection criterion (e.g. kg milk yield), the selection intensity *i* is unit-less (e.g. *S*=2000 kg, σ_P=1000 kg, *i*=2.0). As a rule, values of *i* are in the range from –3 to +3.

Figure 5 gives a graphic illustration of the relation between proportion selected and selection intensity. It is important to remember that the smaller the proportion selected, the better is the selected group, and the higher is the selection intensity

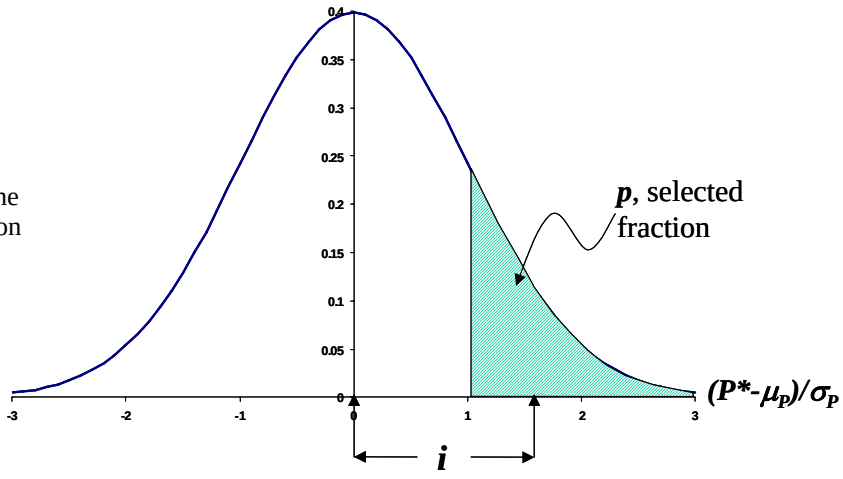


Figure 5. Illustration of the relation between proportion selected (*p*) and selection intensity (*i*).

Calculation of selection intensity from proportion selected, Appendix 1

The values needed for computing *i* for a given proportion selected are available in standard tables for the normal distribution. A small collection of values from such tables are given in the table in appendix 1 with *p* being the top or bottom fraction of the distribution. The same table can also be used for computing *i* when an intermediate fraction (e.g. those between the limits corresponding to 10% and 20%), as will be described in appendix 2.

Selection intensity from intermediate fractions, Appendix 2

Selection intensity in small samples, Appendix 3

The table in appendix 1 is only valid if we have large samples, usually greater than 500 individuals. For smaller samples, see appendix 3.

A more detailed illustration of how the selection intensity is influenced by the proportion of animals that is selected is given in Figure 6. We can see that a low proportion of animals selected will give high selection intensity. The proportion selected can be expressed as:

$$\text{Proportion selected} = \frac{\text{number of animals selected}}{\text{number to choose among (evaluated)}}$$

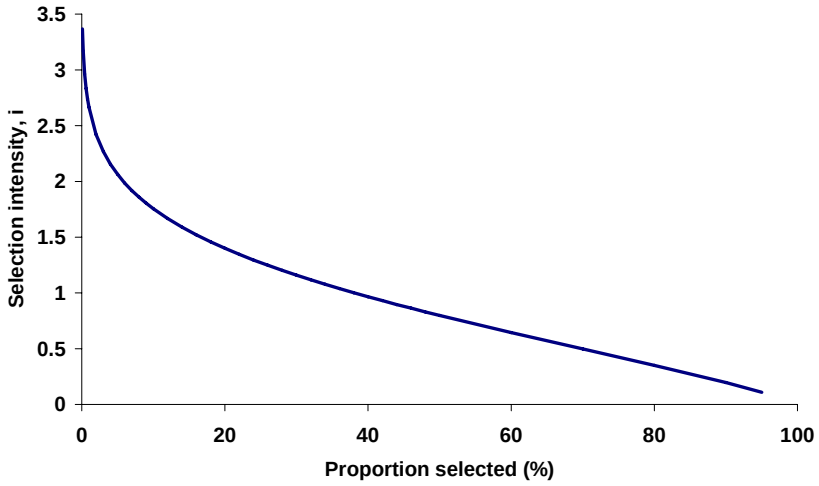


Figure 6. Illustration of the change in selection intensity (*i*) as a result of a change in proportion selected (*p*).

We can keep the proportion selected low by selecting extremely few male animals. This will, however, increase the risk for inbreeding and it might be deleterious if one of the few animals used widely would carry a defect that was not detected when the animals were evaluated. So, we need to be a little careful in selecting too few animals. What can be done without risk, though, is to lower the proportion selected by increasing the number of animals that are evaluated, and thus have more animals to choose among. The relations between selection intensity and genetic response will be discussed later.

General equation for genetic change from truncation selection

Now that we have defined the standardized selection differential, i.e. the selection intensity, we can define a general equation for response to truncation selection.

Let's assume that we have a breeding goal (what we want to improve) defined as:

Breeding goal *(or* ***T****rue breeding value)* $= T = v_1A_1 + v_2A_2 + \ldots + v_mA_m =$ **v'a** **[13]**

with *m* traits and corresponding economic weights. We try to predict this breeding goal using an index:

Index *(or predicted breeding value)* $= \hat{T} = I = b_1X_1 + b_2X_2 + .. + b_nX_n =$ **b'x** *[14]*

based on *n* phenotypic measures, each with a corresponding index weight. We calculate *I* such that the correlation between goal and index (r_{TI}) is maximized.

We have already seen (in equation [11]) that the change to the next generation is:

$$\Delta T_g = \overline{I}_{sel}$$

assuming there is equal selection in both parental sexes. The change, i.e. $\overline{I}_{sel}$, is actually the genetic selection differential (the difference between the parent and

the progeny generations). If we want to express this in terms of selection intensity we need to divide with σ_I (just like we divided the phenotypic selection differential with σ_P, see equation [12]). The selection intensity expressed in index units will then be: $i = \overline{I}_{sel} / \sigma_I$. We can thus express the genetic change to the next generation as:

$$\Delta T_g = \overline{I}_{sel} = i\sigma_I \qquad [15]$$

and we can see that the genetic change will depend on both how intensively the parents are selected and the precision of the genetic evaluation.

The influences of selection intensity and precision on genetic change are illustrated in Figure 7. First you can notice that the variation in predicted breeding values (index) is lower than the variation in phenotypic values. This is because the variation in ***true*** breeding values is only a fraction of the phenotypic variation (h^2) and that the index is not a perfect representation of the true breeding values, as we do not have unlimited information.

With increasing precision of the selection criterion (low precision in the left part of Figure 7 and high precision in the right part), the variation in index values is increasing. Given the same intensity of selection, *increased variation in index values increases the selection response*, e.g. compare the difference between parent and offspring averages in the two upper situations. On the other hand, given the same precision, *increased selection intensity increases selection response*, e.g. compare the difference between parent and offspring averages in the two left situations.

Now, as we have seen in the derivation of selection indexes, the variance of the index is a measure of how precise the index is, the more information we have, the larger the σ_I. An even more convenient measure of the precision of the index is the reliability, i.e.

$$r_{TI}^2 = \frac{\sigma_I^2}{\sigma_T^2} \qquad [16]$$

or the "accuracy":

$$r_{TI} = \frac{\sigma_I}{\sigma_T} \qquad [17]$$

If we solve for σ_I in [17] we find that $\sigma_I = r_{TI} \cdot \sigma_T$. Entering that into [15] gives:

$$\boxed{\Delta T_g = \overline{I}_{sel} = i\sigma_I = i\ r_{TI}\ \sigma_T} \qquad [18]$$

where the last part probably is the most commonly used expression for genetic response to truncation selection.

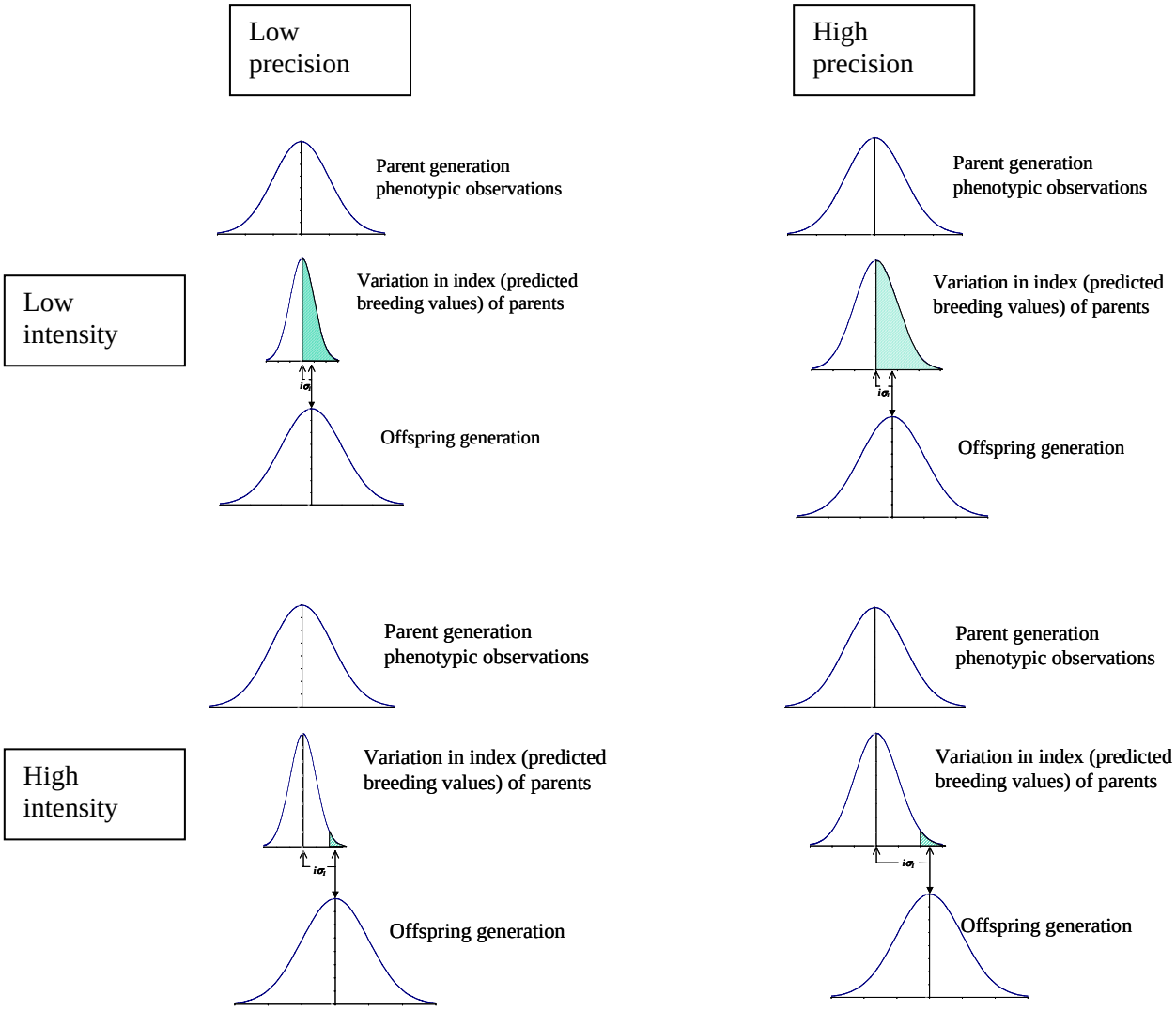


Figure 7. Illustration of the relationship $\Delta T_g = \overline{I}_{sel} = i\sigma_I$.

In the left part the precision of the index is low (e.g. for a trait with low heritability measured on the individual itself or on few progenies), whereas in the right part precision is high (high heritability or many progenies). In the upper part the selection intensity is low (truncation point 0.0, proportion selected 50%) and in the lower part the intensity is high (truncation point 2.0, proportion selection ca 2.2%). Note the effect on difference between parent and offspring generation.

Calculation of response in each breeding goal trait, Appendix 4

Equation [18] describes the genetic change in the total breeding goal. Occasionally it is interesting to get the response in each of the breeding goal traits separately, e.g. to see whether the response in mastitis resistance is favourable or not. This can be calculated as:

$$\Delta T_{gj} = b_{T_j / I}\, i\, \sigma_I = \mathbf{b}'\mathbf{g}_j\, i / \sigma_I \qquad [19]$$

where $\mathbf{g}_j$ is the column vector corresponding to trait *j* in matrix **G** in the selection index equations (**Pb=Gv**). The derivation is described in more detail in appendix 4.

The general equation applied to phenotypic selection

Let's apply the general equation [18] to the simple example of phenotypic selection, i.e. one trait measured on the individuals themselves.
Then:

$$\sigma_T = \sigma_A \text{ (as T = A) and } r_{TI} = \sqrt{h^2} = h$$

which gives:

$$\Delta T_g = i \cdot h \cdot \sigma_A = \frac{S}{\sigma_P} \cdot h \cdot \sigma_A = h^2 \cdot S \qquad [20]$$

which is the same as equation [10].

Summary of the general equation for genetic response per generation with several selection paths

In summary, the selection response per generation, assuming that both sexes are selected with equal intensity is:

$$\Delta T_g = \bar{I}_{sel} = i\sigma_I = i\, r_{TI}\, \sigma_T \qquad [21]$$

If there are different selection intensities (i_S and i_D) or accuracies (r_{TI_S} and r_{TI_D}) in the two selection paths, sires (*S*) and dams (*D*), we instead get:

$$\Delta T_g = \frac{\bar{I}_S + \bar{I}_D}{2} = \frac{i_S\, r_{TI_S}\, \sigma_T + i_D\, r_{TI_D}\, \sigma_T}{2} \qquad [22]$$

In some breeding programs, e.g. in dairy cattle, we may have an even more complicated picture, where we have different selection intensity and accuracy among parents breeding sons and parents breeding daughters, which leads to the following equation:

$$\Delta T_g = \frac{\bar{I}_S + \bar{I}_D}{2}$$

$$= \frac{\left(\frac{\bar{I}_{SS} + \bar{I}_{DS}}{2}\right) + \left(\frac{\bar{I}_{SD} + \bar{I}_{DD}}{2}\right)}{2} \qquad [23]$$

$$= \frac{\bar{I}_{SS} + \bar{I}_{DS} + \bar{I}_{SD} + \bar{I}_{DD}}{4}$$

where the four paths of gene transmission are:

SS = sires to breed sons (the next generation of sires)

DS = dams to breed sons

SD = sires to breed daughters (the next generation of dams)

DD = dams to breed daughters

How to predict genetic change per year

So far, we have only looked at genetic change from one generation to the next, i.e. per generation. If instead we want to find out what change can be expected in, say, 10 years, we need to know the genetic change per time unit, usually per year.

The most intuitive way to get to the genetic change per year is probably to say: OK, I know the genetic change in one generation, so: how many generations are there in one year? The answer is:

a) if there are ***two*** generations produced per year (i.e. each generation takes half a year to complete) then I have twice the genetic change per year compared with the genetic change per generation.

b) if it takes two years for a generation to replace the previous (***half*** a generation per year), I have to multiply the genetic change per generation by 0.5.

So, for the two examples above we would get the genetic change per year as:

$$a) \quad \Delta T_y = \Delta T_g \times 2.0 = \Delta T_g \times \frac{1}{0.5} \qquad [24]$$

$$b) \quad \Delta T_y = \Delta T_g \times 0.5 = \Delta T_g \times \frac{1}{2.0} \qquad [25]$$

where the first calculation in a) and b) is to multiply ΔT_g by the number of generations per year, and the second calculation is to multiply by the inverse of the time between successive generations. Even though the first approach may be easier to understand, it is usually the second that is used in practice (but of course they are identical). The parameter in the denominator is called the ***generation interval***, and is defined as *the average age of the parents at the birth of the offspring that are to be selected as breeding animals*.

If we then write the expected genetic change per year in a general way (assuming that males and females are selected in the same way) we get:

$$\Delta T_y = \frac{\bar{I}_{sel}}{L} = \frac{i\sigma_I}{L} = \frac{i\, r_{TI}\, \sigma_T}{L} \qquad [26]$$

From this we can see that our chances to get a high annual genetic progress would seem to be better if we:

- select for traits with large additive genetic variation (as σ_T depends on σ_A)
- have a high accuracy (r_{TI}) in our genetic evaluation
- have a high selection intensity (i), i.e. select a small proportion of the animals that we choose among
- have a short generation interval (L)

As we will discuss later, the various components that affect the genetic progress are to some extent dependent of each other. Increasing the accuracy may result in lower selection intensity and a longer generation interval. For a maximum genetic progress we need to ***optimise*** these components!

It is not unusual that the age of male and female parents differs, e.g. one might use rams only for a couple of years whereas the ewes produce lambs during 6-8 years. Therefore, we will have 3-4 generations of males within the same time we have one generation of females. In this situation we can describe the genetic change per year as (Dickerson and Hazel, 1944):

$$\Delta T_y = \frac{\left(\frac{\bar{I}_S + \bar{I}_D}{2}\right)}{\frac{L_S + L_D}{2}} = \frac{\bar{I}_S + \bar{I}_D}{L_S + L_D} \qquad [27]$$

Finally, if we have four selection paths with potentially different selection and generation intervals we get (Rendel and Robertson, 1950):

$$\Delta T_y = \frac{\left(\frac{\bar{I}_{SS} + \bar{I}_{DS} + \bar{I}_{SD} + \bar{I}_{DD}}{4}\right)}{\left(\frac{L_{SS} + L_{DS} + L_{SD} + L_{DD}}{4}\right)} = \frac{\bar{I}_{SS} + \bar{I}_{DS} + \bar{I}_{SD} + \bar{I}_{DD}}{L_{SS} + L_{DS} + L_{SD} + L_{DD}} \qquad [28]$$

Genetic gain with different number of progeny within path, Appendix 5

Additional refinements of this expression are needed if all parents in any one of the four paths do not have the same opportunity to produce progeny for breeding (Appendix 5).

The components of genetic gain are inter-related

Let's have a look again at the components of the genetic change (Figure 8):

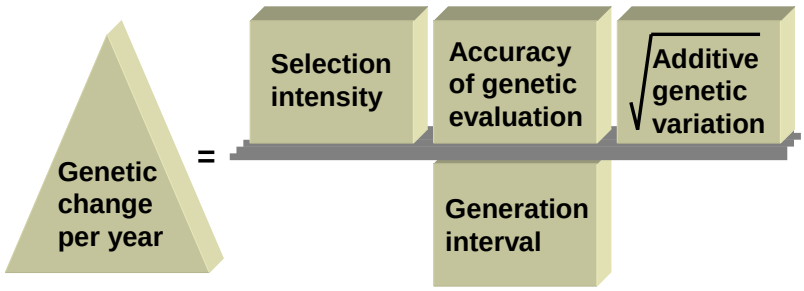


Figure 8. Components of annual genetic change

We can see that there will be no genetic progress, unless there is additive genetic variation in the traits that we want to select for. So, we need to have good estimates of this component, and we need to select in a way so that we don't exhaust the existing additive genetic variation.

As for the other components of the genetic progress, it might seem that if we maximize selection intensity (i) and accuracy (r_{TI}), and minimize the generation interval (L), we would maximize annual genetic progress (assuming that the additive genetic variation in the trait is constant). Unfortunately, life is not that simple. Some of these components are inter-related and trade-offs exist between them. This means that improving one component may lead to a deterioration of another. There are many potential trade-offs that may exist in a certain situation but the most common ones are described in the following.

Increased precision often results in longer generation interval

For selection to be efficient we need to rank the animals on breeding values that are predicted with a fairly good precision. We can increase the accuracy of the genetic evaluation by using observations on progenies, instead of just using an animal's own performance. This means that we have to wait longer for the results, so the generation interval usually increases. How much the increase will be depends largely on the species.

The accuracy can be increased also by increasing the number of progenies per sire. This might increase the generation interval, but not necessarily so. It depends on whether it takes longer time or not to get the results from the additional progenies. Yet another way to increase the accuracy is to use repeated records of a trait. If the records can be done within a short time span the generation interval will not be influenced, but if the repeated record is a second lactation, for example, the generation interval will be increased.

How long the generation interval will be also depends on how early in an animal's life a trait can be measured. Some traits can only be measured late, which means that also performance testing may give long generation intervals. In such situations we may need to select on a correlated trait that is expressed early, even if it means that the accuracy will be lower. For example, the performance of riding horses in competitions can not be evaluated until the horses are 10-12 years old. By predicting the breeding values from simplified performance tests at 3-4 years of age the generation interval will be shortened, but the accuracy of the breeding values will be reduced. Another example might be selection just on pedigree records. This will result in the lowest possible generation interval, but it will also give a low accuracy. This can, however, be improved if a second selection is carried through once the animals themselves have got records.

What precision (accuracy) is required largely depends on how intensively the selected animals will be used in breeding. Some AI-bulls, for example, will be used very intensively, and a very high precision then has to be required.

Increased precision may lower selection intensity

Increasing the precision might give the result that fewer animals can be tested, and thus lower the selection intensity. If animals are performance tested on-farm it will be possible to test a large number of animals and the selection intensity will be high. We may want to increase the accuracy by testing them on-station instead, which will most certainly reduce the number of tested animals and lower the selection intensity.

Our testing capacity is usually limited. For example, we can increase accuracy by increasing the number of progeny per sire, but then each sire will use a larger proportion of the cows available for mating, and the result is that fewer sires can be tested.

Generation interval and selection intensity

The two components generation interval and selection intensity are also interrelated. For instance, a long female generation interval leads to lower replacement rate and higher intensity, as very few animals need to be replaced. But there are also examples of an opposite relationship, e.g. selecting only among young animals gives a short generation interval, but also fewer animals to choose among, compared to selecting among animals of all ages.

In conclusion, improving one of the components influencing genetic change does not necessarily lead to a higher genetic progress. What we should do is to optimize the components!

Prediction of correlated genetic change

Selection on one trait may result in a correlated change in another trait if the two traits are genetically correlated. If the genetic correlation is *favourable* then we can expect a ***favourable correlated genetic change*** in the other trait. The relationship between the traits can be utilized for ***indirect selection***. This is actually what we do when we measure trait 1 (e.g. backfat thickness) but include trait 2 (e.g. meat %) in the breeding goal. The genetic correlation between the two traits must be fairly high for indirect selection to be effective. Assuming the genetic correlation between the traits this high, indirect selection might be motivated if:

trait 1 (measured)	***trait 2 (in breeding goal)***
has a high h^2	has a low h^2
can be measured on live animals	is sex limited or is a carcass trait
is expressed early	is expressed late in the animal's life
is cheap to measure	is expensive to measure

If, on the other hand, a trait that we select for shows an *unfavourable genetic correlation* with another trait, then we can expect an ***unfavourable correlated genetic change*** in that other trait, unless we counteract this by including both traits in our selection. For example, selection for increased egg number in poultry will result in reduced egg weight, unless both traits are considered. In the same way we can predict that selection for increased milk production in cows will result in reduced resistance to mastitis. Both these traits are included, therefore, in the bull index in many countries. Already quite low genetic correlations may result in unfavourable correlated changes. So, before starting up a selection program, estimate genetic correlations between all traits of importance and predict what changes can be expected also in traits not planned to be included in the selection.

So, how can we predict the correlated change in a trait (e.g. trait 2) when we select on another trait (e.g. trait 1)? Let's start by calculating the genetic change in trait 1 when selecting on an index based on this trait. As we have seen before the genetic change in trait 1 is:

$$\Delta T_1 = i\, r_{TI}\, \sigma_{T_1} \qquad [29]$$

The correlated genetic change in trait 2 when selecting on trait 1 ($\Delta T_{2/I_1}$) can be calculated by multiplying the genetic change in trait 1 by the genetic regression of trait 2 on trait 1 ($b_{T_2|T_1}$):

$$\begin{aligned} \Delta T_{2/I_1} &= b_{T_2|T_1} \times \Delta T_1 = \frac{\sigma_{T_1T_2}}{\sigma^2_{T_1}} \times \Delta T_1 \\ &= \frac{\sigma_{T_1T_2}}{\sigma^2_{T_1}}\, i\, r_{TI}\, \sigma_{T_1} = \frac{\sigma_{T_1T_2}}{\sigma_{T_1}\sigma_{T_2}}\, i\, r_{TI}\, \sigma_{T_2} = r_{g_{12}}\, i\, r_{TI}\, \sigma_{T_2} \end{aligned} \qquad [30]$$

In conclusion, the correlated genetic change in trait 2 when selecting on trait 1 is:

$$\boxed{\Delta T_{2/I_1} = r_{g_{12}}\, i\, r_{TI}\, \sigma_{T_2}} \qquad [31]$$

where

$r_{g_{12}}$ is the genetic correlation between the two traits

i is the selection intensity

r_{TI} is the accuracy when trait 1 is the selection criterion

σ_{T_2} is the additive genetic standard deviation for the breeding goal trait which acts as a scaling factor to express the correlated change in units of trait 2.

Correlated genetic change is illustrated in Figure 9:

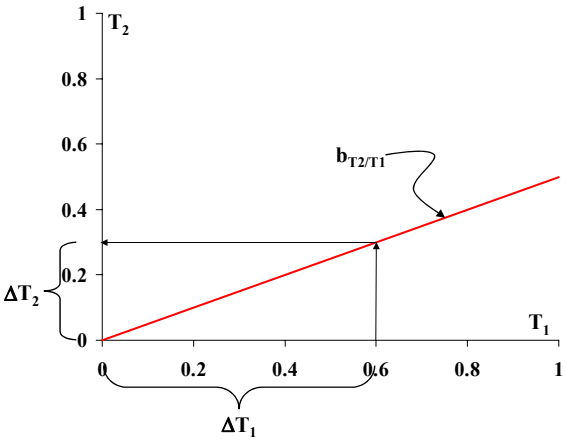


Figure 9. Illustration of the correlated genetic change in trait 2 when selecting on trait 1.

Efficiency of indirect vs direct selection, Appendix 6

Sometimes it may be more efficient to select indirectly for a trait one wants to improve, rather than to select directly for it. This can occur when the heritability of the correlated trait (e.g. somatic cell count in dairy cattle) has a higher heritability than the trait we want to improve (mastitis resistance) and the correlation between the two traits is high. For a more detailed description of the efficiency of indirect selection, see appendix 6.

Realized genetic change

When a breeding program has been in operation for some years it should always be checked what genetic changes that have been achieved. This is the ***realized genetic change***, which is not the same thing as the predicted genetic change. Predicting the expected genetic change helps us to optimize a breeding program, whereas estimating the realized genetic change helps us to find out whether the program has been successful. The realized genetic change, often also called the genetic trend, should be checked at regular intervals.

We usually can measure easily the phenotypic change of a trait over time, but that includes both genetic change and changes due to management and environment. So, we need to distinguish the genetic part.

A good and cheap method for estimating the realized genetic change in a trait (or combination of traits**)** is to ***compare BLUP breeding values of animals born in different years***. BLUP breeding values could be plotted against the animals' birth year in a graph (Figure 10). The change in average breeding value over time reflects the realized genetic change.

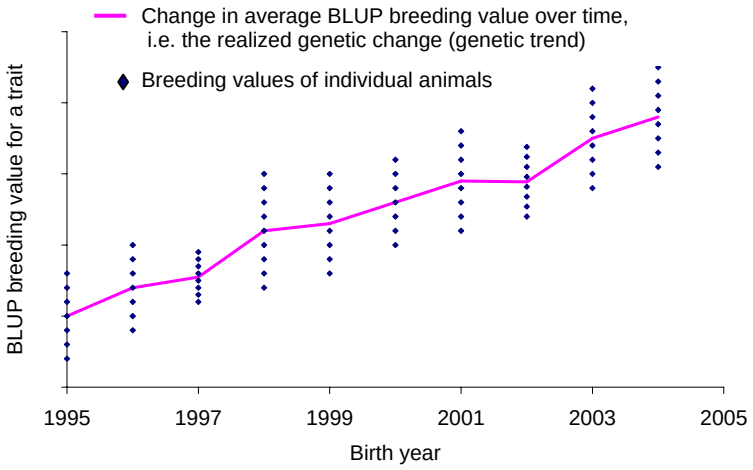


Figure 10. Estimation of realized genetic change by plotting BLUP breeding values against birth year.

A method sometimes used for estimation of realized genetic change is to ***reconstruct earlier populations*** by using the same frozen semen or embryos at two (or more) time events. If semen from the same male animals is used at two points in time, then the male genotype is constant over time and any difference observed in progeny means between the two points is due to genetic change in females and/or changes in environment. Subtracting the change observed in progenies of "constant sires" from the phenotypic change of the total population in the same time period gives:

$$\Delta P_{Total} - \Delta P_{progenies\ \ from\ "old"\ semen} = (\Delta T_{\text{Real}} + \Delta E) - (\frac{1}{2}\Delta T_{\text{Real}} + \Delta E) = \frac{1}{2}\Delta T_{\text{Real}}$$

and by multiplying the above phenotypic difference by 2 we get an estimate of the realized genetic change. If embryos are used instead of semen, then the full genotype is kept constant over time, and any change observed in this part of the population will be due to environmental change. Subtracting the phenotypic change in the embryo progenies from the total phenotypic change will give directly the estimate of the realized genetic change.

A drawback of the method of reconstructing earlier populations is that it requires specific "matings" and that using animals in breeding that were selected a number of years ago has a lower economic value (assuming there is a continuous genetic improvement of the population).

Yet another method for prediction of realized genetic change is to keep an ***unselected control population*** for comparison with the selected population. This might be used in selection experiments, but hardly in practical animal breeding. Furthermore, unless the control population is quite big, there is a risk that genetic drift will cause changes of gene frequencies in this population.

The realized genetic progress in a trait is often lower than the change predicted when the breeding program was planned. This might be due to errors in the genetic parameters or in the breeding values, or there might be varying amount of information for different traits in combination with strong and unfavourable genetic correlations between the traits. It could also be that more animals are selected in practice than planned, or that the generation intervals are longer than expected. The breeding program might still be the most optimal one, but if there are big discrepancies between expected and realized genetic change, then a thorough analysis of the program should be performed and revision made accordingly. There should also be a check of the rate of inbreeding.

Selection and genetic variation

So far, we have studied the response to selection implicitly assuming that the genetic variances and covariances (or heritabilities and genetic correlations) are known and constant over time. This is naturally a simplification. In this section we will see what happens with genetic variation and correlations as a result of selection itself, and how that in turn affects the response to selection.

From population genetics we know that selection for a trait determined by one locus will lead to one of the alleles in that locus becoming fixed and the other allele disappearing from the population. When that has happened there is no genetic variation left and there will be no response to selection. Now, it is quite obvious that in nature (and agriculture) there is still a lot of genetic variation around and we can see selection response in the traits selected for. So why hasn't all genetic variation disappeared?

More details on selection and change in genetic (co)variances, Appendix 7

In order to have a meaningful discussion we first have to distinguish between what kind of genetic model we assume. The two main models are called the *finite locus model* and the *infinitesimal model*. For a detailed description of these models and how the genetic variances and covariances are expected to be affected by selection, read appendix 7. Here in the main text we will only give the main results from that appendix.

In the *finite locus model*, where we assume a few loci affecting the trait, selection is generally expected to deplete genetic variation by fixing all favourable alleles. However, there are several points to be remembered.

- During the course of selection the genetic variance might actually increase, depending on the gene frequencies.
- Selection for a dominant allele can be very slow at high frequencies of the favourable allele, and fixation is reached only very slowly.
- Overdominance could contribute to non-fixation.
- If selection is for increased value of two traits, the expectation is that the genetic covariance (correlation) moves towards negative values.
- Antagonistic covariances among subtraits involved in a bigger trait complex (e.g. fitness) may lead to a selection plateau, even if there still exists genetic variation in the subtraits.
- Mutation can create substantial genetic variation, especially in large populations, and can counterbalance loss of genetic variation due to selection and drift.
- Migration can also counteract the forces of drift and selection.

In the *infinitesimal model* we assume an infinite number of genes, all with identical and infinitesimal (very small) effects. Therefore, selection produces no gene frequency changes, or at least they are so small that they can be ignored. In this model the genetic variance consist of the variance contributed from each locus, and the covariances between all loci:

$$\sigma_G^2 = \sigma_{G1}^2 + \sigma_{G2}^2 + \ldots + \sigma_{G12} + \sigma_{G13} + \ldots \qquad [32]$$

Linkage disequilibrium, for more details see Appendix 8

In a random mating population there is no association between the genotypes at different loci (i.e. no covariances), which means that the genetic variance is just the sum of the variance components. With selection for increased genotypic value we will, however, create linkage disequilibrium between loci, which will create negative covariances between loci. This is called the ***Bulmer effect***. It means that the genetic variance (σ_G^2 in [32]) will be lower than at random mating. Similarly, it can be shown, that if we select for increased value of two traits, the genetic covariance and the genetic correlation between the two traits will move in the negative direction.

So, the genetic variance σ_G^2 and the genetic covariance between traits decrease also in the infinitesimal model, as a result of selection. However, the difference compared with the finite locus model is that the change occurs quickly (basically in the first two generations of selection) and that the variance and covariance returns to the initial value after selection ceases.

It is quite obvious that neither of the two described models is in any strict meaning true. Of course there cannot be an infinite number of genes, each having equal effects. We know that there are so-called major genes or QTLs out there. On the other hand, the finite locus model predicts a depletion of genetic variance with selection that we do not see in real populations.

It is more useful to consider these two models exactly as that: *models*. There are certain features of both models that can be found in real data, and where predictions from the models can be useful. For instance:

- A decrease in genetic variance with selection due to the Bulmer effect can be seen in random mating populations that undergo selection. At the same time, however, there are most likely also gene frequency changes occurring, at least for some loci with large effects on the trait.
- The genetic correlation between two traits can be expected to move in the negative direction, either due to the Bulmer effect or due to fixation of pleiotropically favorable alleles.
- Genetic variance is not depleted quickly which may be explained by:
  - there are many genes affecting the trait
  - mutations may play a role
  - gene effects may change over time or over environments (genotype by environment interaction)

General results from long term selection experiments

One way to test the theoretical models is to actually carry out the selection in specially designed experiments. For obvious reasons, these are usually not done in animals with long generation intervals, such as dairy cattle, but for general questions about genetic models mice and fruit flies, or even plants, are good enough.

Even though there is a large variation in the outcome of long-term selection experiments, some general conclusions can be drawn (Walsh and Lynch, 2000):

- The mean of the population at the end of selection is usually far outside the range in the base population.
- Response can be very uneven (periods of increased response often found)
- Additive genetic variance can increase during parts of a selection experiment.
- Reproductive fitness usually declines when selecting on a single trait
- Most populations approach an apparent selection limit
- Considerable genetic variation exists at the selection limit

We will exemplify all these points with examples from some selection experiments. First we will look at an experiment (Yoo, 1980) in which selection was for increased number of abdominal bristles in *Drosophila melanogaster for 90 generations* (Figure 11). Note that during most of the time the genetic response is almost linear, but for some periods (around generation 50 and 75) the selection response jumps quite drastically. This might be because the linkage disequilibrium that was created from the selection somehow was broken, releasing new genetic variation that could be exploited by selection. Another reason could be that rare favorable alleles increased in frequency with selection, thus increasing the genetic variance and the response (cf. Figure 4 in appendix).

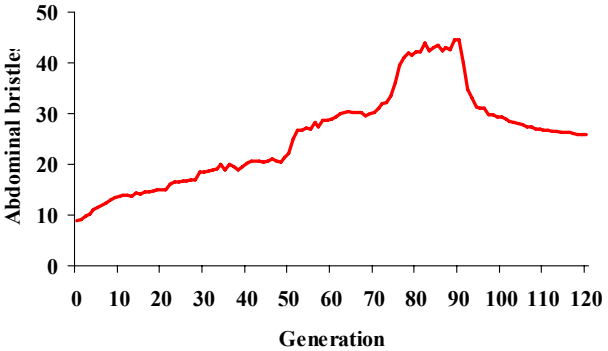


Figure 11. Change in abdominal bristle number after selection for 90 generations and random mating for 30 generations (Yoo, 1980).

At the end of the 90 generation period, it seems that the population has reached a *selection limit*. Selection was stopped after generation 90, but the line with random mating was kept for another 30 generations (Figure 11). There was a fast decline in bristle number as a result of the discontinued selection. This indicates that the selection plateau was due to an antagonism between bristle number and fitness in the last generations before generation 90. Therefore, it was not possible to achieve further upward response in bristle number. The decrease in bristle number when selection ceased could be a result of natural selection favouring fitness rather than bristle number. The decrease in bristle number also shows that the lack of selection response before was not due to lack of genetic variation.

In another famous selection experiment, the Illinois maize experiment (Dudley, 1977), which started in 1896 (!), selection has been for increased (and decreased) oil content in the seeds (Figure 12). The upward selection line did not reach a selection limit even after 75 (or even 90, not shown) generations. The genetic level at generation 90 was 22 genetic standard deviations higher than the mean of the base population, i.e. way outside the range in the base population. The downward line *did* reach a limit, but a natural one – oil content cannot go below zero. That this selection limit was not due to lack of additive genetic variation could be seen when the downward selected line was split and selected upwards in one line instead – an increase in oil content was seen (Figure 12). Selecting downwards in the upwards selected line also produced a response, but that was not unexpected, because there was still response also in the upward direction.

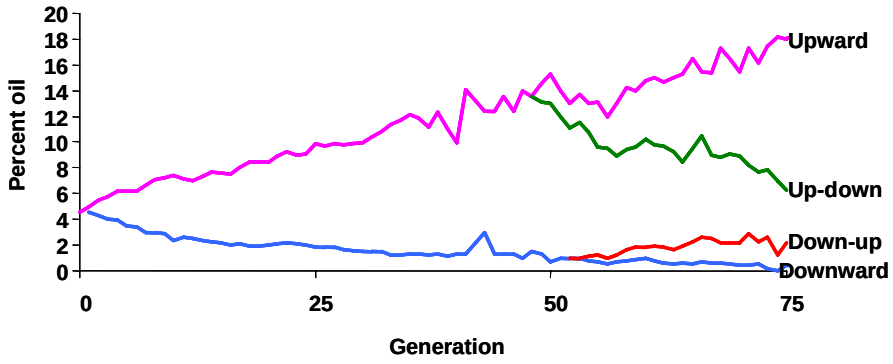


***Figure 12**. Change in oil content of maize from selection upwards and downwards. Around generation 50 also selection in the opposite direction in each line. Adapted from Dudley (1977).*

Concluding remarks

Given the results from selection experiments, and the expectations from the two genetic models, it is quite clear that the prediction from the finite locus model regarding the depletion of genetic variance with selection does not seem to be correct. Even after many generations, there is still ample genetic variance left. When selection limits seem to occur, they usually do not seem to be due to lack of genetic variation. Other constraints, such as linkage disequilibrium or antagonistic correlations with other traits (including fitness), seem more important. These constraints might, at least partly, have been created by selection itself.

Rather than assuming that a quantitative trait is influenced by a very large number of genes all with very small and identical effects, it makes more sense to assume that a few genes have rather large effects on the trait, some other genes have a moderate effect, but that there is also a large number of genes each having a small effect. This can be thought of as an L-shaped or exponentially shaped distribution (Figure 13).

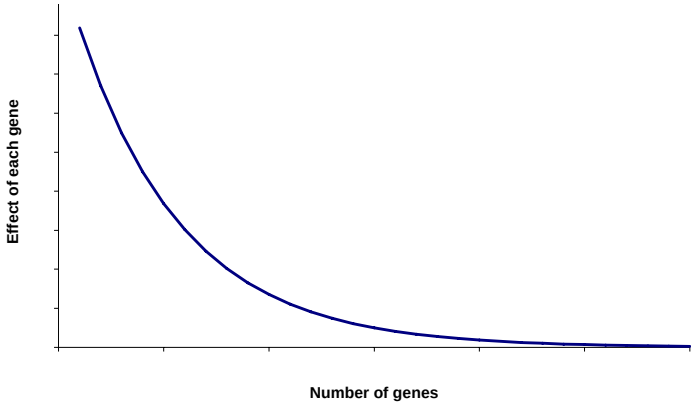


Figure 13. Illustration of a genetic model with a few genes with large effects and many genes with small effects

With this model we would expect the genes in the left part of the graph to behave according to the finite locus model and the others more according to the infinitesimal model. This model combines the features of the two models discussed previously, and may be a better representation of reality, although still – it is only a model.

When it comes to animal breeding, most traits seem to be closer to the infinitesimal model than to the finite model, and there is little evidence of decreased genetic variation due to selection. One reason could be that mutations create new genetic variation which can counterbalance loss of genetic variation due to selection and drift. Another important factor is that selection often includes both production and fitness traits. Furthermore, if there is an interaction between genotype and environment, changes in the environment over time will also change the genetic variance. The environment will influence to what extent various genes are expressed.

So, the limitation does not seem to be depleted genetic variation. A more severe limitation seems to be unfavourable genetic correlations between traits that we want to improve. Although these correlations, at least partly, may be the result of selection for several traits simultaneously, there is no other way to go than to include in selection all traits that describe a well-functioning animal.

References

Dickerson, G. E. and L. N. Hazel. 1944. Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. *J. Agric. Res.* 69:459

Dudley, J. W. 1977. 76 generations of selection for oil and protein percentage in maize. Pages 459-473 *in* Proc. Int. Conf. Quant. Genet. E. Pollak, O. Kempthorne, and T. B. Bailey, eds. Iowa State Univ., Ames, Iowa, USA.

Rendel, J. M. and A. Robertson. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J. Genetics*. 50:1-8

Walsh, B. and M. Lynch. 2000. Chapter 16. Long-term response: Deterministic aspects. *In:* Evolution and Selection of Quantitative Traits. http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html

Some literature for further reading

Bourdon, R.M.1997. *Understanding Animal Breeding*. Prentice-Hall. Upper Saddle River, New Jersey, USA

Falconer, D.S. and T.F.C. Mackay. 1996. *Introduction to quantitative genetics*, 4th ed. Longman Group, Essex, UK

Walsh, B. and M. Lynch. 2000. *Evolution and Selection of Quantitative Traits*. http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html