A Practical guide to genetic gain

Jessica E. Rutkoski[a,*]

[a]*International Rice Research Institute, Rice Breeding Platform, DAPO Box 7777 Metro Manila 1301, Philippines*

Abstract

An understanding of the inheritance of quantitative traits, those with a continuous phenotype, was first established in the early 1900s. This was instrumental for breeding because quantitative genetic theory provides the basis for the development of methods which can be used to increase the rate of genetic improvement, referred to as "genetic gain", within a breeding population over time. Today, the concept of genetic gain and its basis in quantitative genetics is often not well understood among crop breeders and scientists, often resulting in inefficient or ineffective crop improvement efforts. This chapter aims to provide clarity on genetic gain to help those engaged in crop improvement to take actions that will enable them to be more successful. To do so, a thorough introduction to genetic gain and the population improvement cycle is provided along with a review of selection techniques essential for breeding. Next, I demonstrate why genetic improvement on a population basis is needed to facilitate variety development. In order to show that the genetic gain is tractable, the theory behind genetic gain and its prediction is explained, followed by a discussion on realized genetic gain including a review of methods that can be used for its estimation. Lastly, guidance is given on how to improve rates of genetic gain in applied breeding programs.

Keywords: Genetic gain, plant breeding

1. Introduction

For the majority of agricultural history, selective breeding of animal and plant species was done without a formal understanding of how selection leads

*Corresponding author
Email address: j.rutkoski@irri.org (Jessica E. Rutkoski)

to genetic improvement. The basic principles of inheritance were first described by Gregor Mendel (1866) and these principles could only fully explain the inheritance of traits that fall into discrete categories. Important characters like reproductive fitenss in natural populations, as well as the yield of grain crops and carcass weight of livestock were known to exhibit a continuous range of variation. These characters are referred to as "quantitaive traits". It wasn't until 1918 when the model for the inheritance of quantitative traits based multiple Mendelain factors was comprehensively described by the statistician and geneticist R. A. Fisher (1918). Also around that time, the geneticist Sewall Wright described many quantitative genetic principles that became instrumental for breeding (Wright, 1920, 1921). Animal breeder, J.L Lush was influenced by Wright's work, and was one of the first to apply quantitative genetics to breeding. The now widely recognized breeder's equation which describes how genetic improvement can be predicted from one generation of selection to the next was just one of Lush's many important contributions (Lush, 1937).

Although ideas about the application of quantitative genetics for breeding are over 80 years old, many plant breeding programs have not yet taken full advantage of this knowledge to increase rates of genetic improvement. This may be because there has been a great deal of emphasis on understanding what are the genes or genomic regions which affect traits, in hopes that this will enable a precise stacking of favorable genes. While such an approach may be useful for monogenic or even oligoenic traits, it is not a solution for the improvement for quantitative traits (Bernardo, 2008) which are known to be influenced by a large number of genes. For such traits, genetic improvement over cycles of selection is necessary so that favorable alleles can be brought together gradually. Such a population improvement approach is how natural selection works to drive adaptive evolution (Fisher, 1930), and it is required for achieving genetic gain in breeding programs. This chapter aims to help improve the understanding and appreciation of this concept for plant breeding.

2. The Basics of genetic gain and population improvement

2.1. Definition

Genetic gain from selection, or simply "genetic gain", is defined as the improvement in average genetic value in a population or the improvement in average phenotypic value due to selection within a population over cycles

of breeding (Hazel and Lush, 1942). Genetic gain may also be referred to as response to selection which may be a better term for describing changes that are not necessarily favorable. In this article, the symbol R is used to represent genetic gain.

Discussions on genetic gain may be focused either on expected genetic gain or realized genetic gain. Expected genetic gain is a prediction of the actual change in phenotype that would occur due to the genetic changes brought about by a proposed selection or a proposed breeding strategy. Expected genetic gain can be estimated using parameters obtainable from breeding experiments and given various assumptions. Realized genetic gain is the observed gain due to selection over cycles. Occasionally, the term genetic gain is used incorrectly to describe phenotypic differences between individuals, such as when describing the yield advantage of a promising new variety compared to a control. It is also incorrect to use the term genetic gain to describe a trend in phenotype over time that was not the result of cycles of selection within a population.

2.2. How genetic gain is achieved

Genetic gain can be achieved for virtually any trait as long as it is heritable. It can also be achieved for total net merit, which is a combination of multiple economically important traits. Selection for net merit is done using an economic selection index (Hazel, 1943; Smith, 1936) which is a linear combination of different traits of economic importance weighted optimally so that selection based on the index maximizes expected genetic gain in net merit. Because the term genetic gain only applies to changes in population mean over cycles of selection, in makes sense to talk about genetic gain in traits that are heritable and conferred by multiple loci. This is because improving such traits can be done most effectively over multiple cycles of selection.

Explicitly stacking favorable alleles to achieve a desired genotype, backcrossing for the introgression of favorable alleles, and mutation breeding are breeding methods that do *not* produce genetic gain because these breeding approaches do not lead to improvement the breeding population as a whole. Even if these approaches were applied population-wide, once the targeted set of desired loci are fixed, there can be no further improvement in genetic value.

Strictly speaking, genetic gain can only be realized from executing at least one cycle of breeding. The breeding cycle required to realize genetic gain, illustrated in **Figure 1**, consists of four parts, generation, evaluation,

selection, and recombination. This process is also referred to as "population improvement" or "recurrent selection". Different breeding materials, such as single non-inbred plants, or families such as inbred lines, may be used for the processes of evaluation, selection, and recombination. For example, an individual's progeny from open-pollination may be used for evaluation and its progeny from self-pollination for recombination (Goulas and Lonnquist, 1976). The breeding materials that are subject to selection are called selection units, those that are evaluated are called evaluation units, and those that are recombined are called recombination units. In self-pollinated crops, the same inbred lines are often utilized as the evaluation, selection, and recombination units.

Generation is the development of the breeding materials that will be used in the evaluation, selection, and recombination processes. The time required to complete this step of the breeding cycle can be reduced by using off-season nurseries or greenhouses (Watson et al., 2018; Collard et al., 2017), or by using doubled haploid technology (Gallais and Bordes, 2007) when inbred lines are required.

Evaluation is the collection of phenotypic and/or genotypic data that will be used in the selection process and it is the generally the most expensive step in the breeding cycle. For phenotypic evaluation, conducting field trials in multiple environments with appropriate field plot technique is a critical component. High-throughput phenotyping and partially replicated experimental designs can help to increase the efficiency of phenotypic evaluation (Cullis et al., 2006; Williams et al., 2010; Haghighattalab et al., 2016), thereby enabling evaluation of a larger number of breeding materials.

Selection is the identification of which selection units to recombine based on the selection criteria which may be single phenotypic values, or multiple phenotypic values combined using a selection index, Best Linear Unbiased Prediction (BLUP) (Henderson, 1975) reviewed by Piepho et al. (2008) or another method for predicting the selection units' value for breeding purposes such as thoes described by Gianola (2013). The selection units are then ranked based on the selection criteria and all those above or below a certain threshold are selected. Selection done this manner is called truncation selection. In truncation selection, there is no attempt made to control how much each selection unit contributes to the next generation nor to control the exact cross combinations made. Optimum contribution selection (Meuwissen, 1997) is an alternative to truncation selection which takes into account the selection criteria as well as the relationships between the selection units

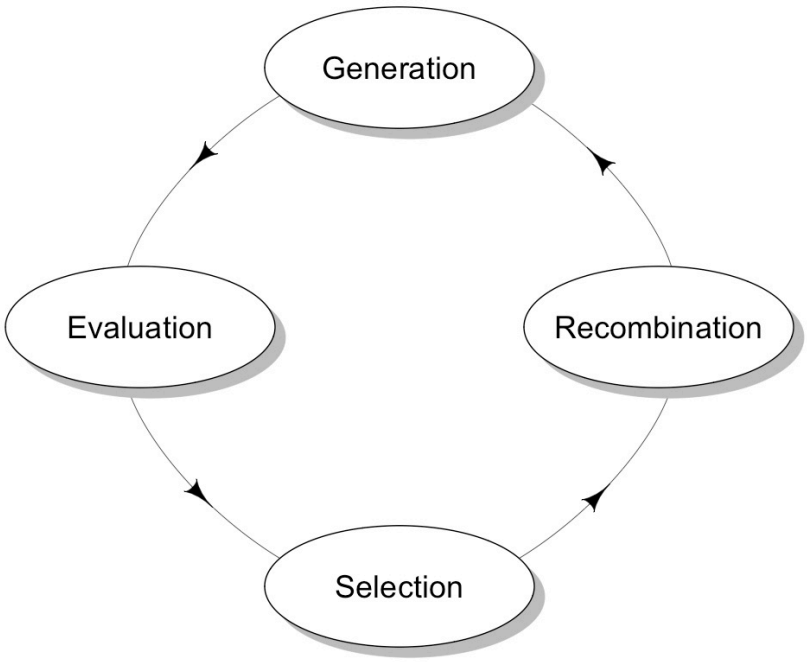

Figure 1: The Breeding cycle. Recurrent cycles of breeding which consist of generation, evaluation, selection, and recombination, are required to achieve genetic gain over time.

to determine how much each one should contribute to the next generation to control the rate of inbreeding. This allows one to control short and long-term genetic gain.

Recombination is the reshuffling of allelic combinations found in the selected breeding materials. This is done by open-pollination or controlled crossing. If crosses are controlled it is possible to precisely plan which pairs to cross, referred to as mate selection or mate allocation. Mate selection is useful for increasing genetic gain for an index of multiple economically important traits if at least one of the traits contributes to economic return in a non-linear way (Allaire, 1980). An example of a trait that would have a non-linear contribution to economic return would be days to flowering where intermediate values are favored as opposed to maximum or minimum values.

Within the breeding cycle, selection may be done in stages. For example, selected bulk and pedigree breeding methods involve evaluation and selection during multiple self-pollination generations (Weber, 1984). With these breeding methods, the breeding cycle consists of additional generate-evaluate-select steps prior to recombination.

Implementation of Genomic Selection (GS)(Meuwissen et al., 2001) can reduce the length of the breeding cycle because the selection units can be single plants that are evaluated by genome-wide genotyping and then selected based on predicted breeding values, where breeding value is the selection unit's value as a parent. This can dramatically hasten the generation and evaluation steps. For example, in self-pollinated crops, single non-inbred plants can be genotyped, selected, and recombined without first deriving inbred lines from these plants and evaluating them phenotypically. However, with GS, generation of evaluation units suitable for phenotyping, and then phenotypic evaluation of these materials is still eventually required in order to generate data that will be used to predict breeding values in future breeding cycles.

2.3. An Example of genetic gain over cycles of breeding

As a specific example of genetic gain realized from a simple breeding program of a cross-pollinated crop, consider a large randomly mating population of maize (*Zea mays*) that is variable in resistance to Northern corn leaf blight (NLB) caused by *Exserohilum turcicum*. Resistance to NLB is a quantitative trait (Poland et al., 2011) and more resistant maize plants are desired. The maize population has not yet been subject to selection and pedigree relationships between individuals are not known. This population, referred to as

the base population, is then used as the initial population of a new breeding program for NLB resistance. To start the first breeding cycle, 500 plants from the base population are grown (generation) and then phenotypically evaluated for resistance to NLB (evaluation). Based on these measurements the breeder selects 50 plants with the highest levels of NLB resistance after flowering (selection), and the selected plants open-pollinate such that the male parent is randomly selected from the population (recombination). The progeny of these crosses forms a new population referred to as the "selected population" from cycle one of selection.

For the second cycle of selection, the generate-evaluate-select-recombine process (**Figure 1**) is repeated, but this time starting from the selected population that was just created. This population is evaluated, the best 50 individuals are identified and open-pollinated, and their progeny become the selected population resulting from cycle two. The breeding cycle is repeated many more times, each time using the selected population that was just developed in the previous cycle in the generate-evaluate-select-recombine process (**Figure 1**).

The genetic gain in NLB resistance realized by the example maize breeding program after each cycle of selection can be observed by comparing the average level of NLB resistance within the selected populations with the average level of NLB resistance of base population as shown in (**Figure 2**). (Assume that resistance is scored such that larger values mean higher levels of resistance and lower values mean lower levels of resistance). For instance, the mean level of NLB resistance of the selected population after one cycle minus the mean level of NLB resistance of the base population is equal to the genetic gain in disease resistance due to one cycle of selection. The observed increase in NLB resistance in the selected population is solely due to heritable genetic factors. The observed differences cannot be due to environment effects because both the base population and the selected population are observed alongside one another in the same environment. As more cycles of selection are conducted, the difference between the selected population mean and the base population mean increases (**Figure 2**).

2.4. Rationale for population improvement

Improvement of a population over cycles of breeding within a mostly closed gene pool is necessary in order to be able to develop a steady stream of varieties that are superior to those released previously. Although it is possible to identify superior new varieties by sampling from a population that

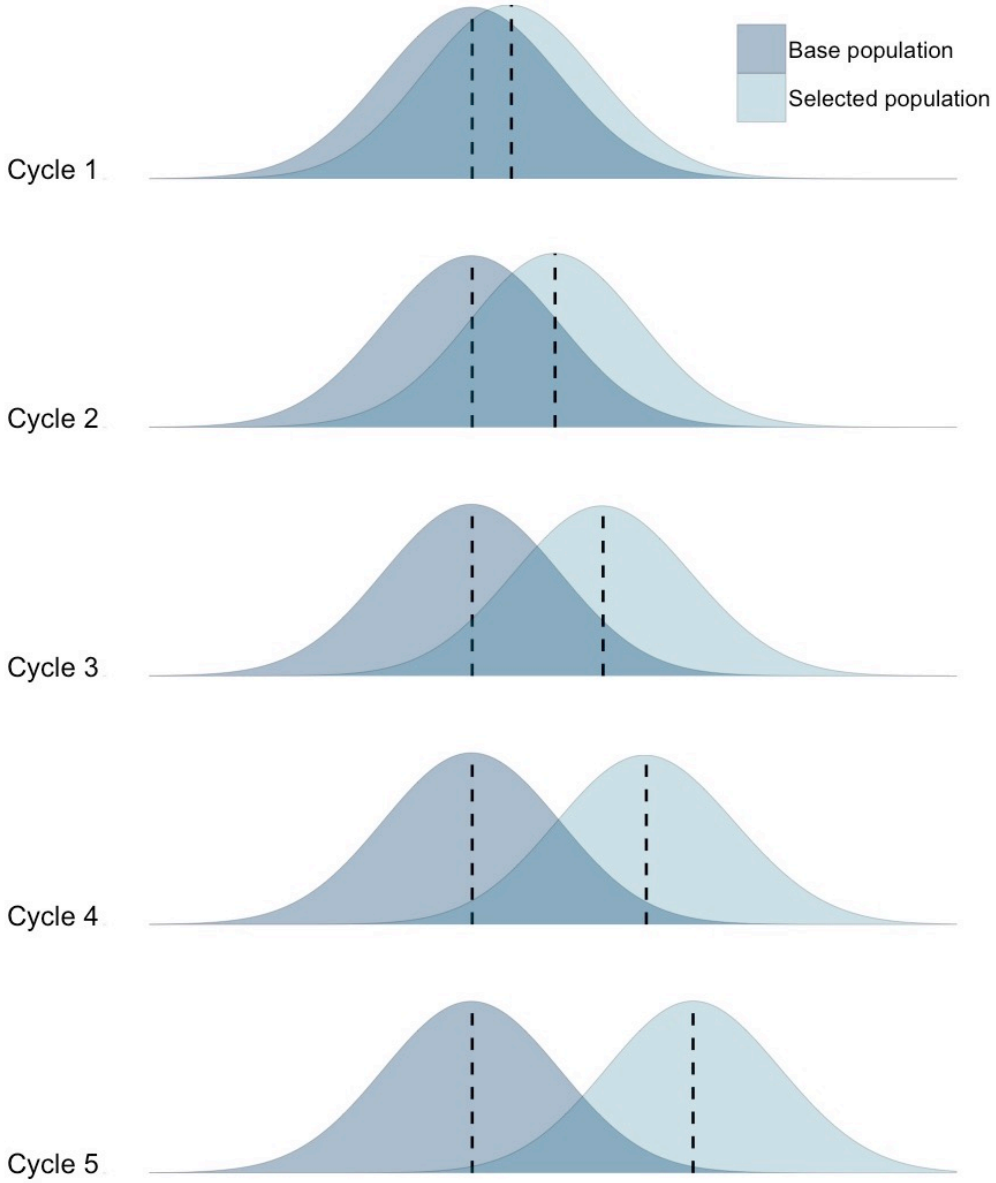

Figure 2: An example of realized genetic gain over five cycles of selection. Histograms depict distributions of phenotypic values for the base and selected populations. A vertical dotted line is drawn at the mean of each distribution. With each advancing cycle the difference between the base and selected population means increases by a small increment. By cycle five, the base and selected population means are significantly different in this example.

is not improving due to ongoing cycles of selection, such as a germplasm collection, achieving genetic gain over cycles of selection increases the probability of identifying superior new varieties on an ongoing basis. Consider for example a large and diverse population from which promising new varieties can be identified. A sample from this population is evaluated and then the best individual is identified for promotion as a variety. Another sample is taken from this population for evaluation and then the best individual is again identified for promotion as a variety. The probability that a superior individual can be identified the second, third, fourth and fifth time in a row is 1/2, 1/6, 1/24 and 1/120 respectively. Thus, by selecting individuals from the same population there is some chance that superior individuals can be be identified each time, but the probability becomes increasingly smaller the more times one attempts to identify an individual that is superior to all others previously identified.

Now consider that instead of sampling from the same static population each time, the population is subject to recurrent cycles of breeding and the best individual is sampled from each successive breeding cycle for promotion as a variety. The probability that a superior individual can be identified each time now partially depends on the genetic gain achieved from each selection cycle. Using a simple simulation carried out in the R language and environment (R Development Core Team, 2016), described in the appendix, the probabilities of successfully identifying a superior variety two to five times in a row are estimated for different levels of narrow sense heritability, h^2, (**Table 1**) assuming a population size of 1000 and that 30 individuals are selected for intermating at each cycle. The parameter h^2 is the squared accuracy of selection and indicates how effective selection will be for achieving genetic gain. The case of $h^2 = 0$ is equivalent to making repeated selections from the same population that is not undergoing selection.

Assuming the heritability ranges from 0.1 to 0.5, the probability that a superior variety is identified five times in a row from a population improving over cycles of selection was shown to be 19 to 87 times greater than the probability of doing so without population improvement. Through this exercise, it becomes clear that breeding is not merely a game of chance. Achieving genetic gain through recurrent cycles of breeding can stack the deck in ones favor by greatly increasing the probability of identifying superior new varieties on an ongoing basis.

Table 1: Probability that an individual superior to all others can be identified every cycle of selection during population improvement. $h^2 = 0$ is equivalent to no population improvement.

	Number of attempts			
h^2	2	3	4	5
0	0.5	0.17	0.04	≈ 0
0.1	0.59	0.33	0.21	0.16
0.3	0.67	0.51	0.45	0.43
0.5	0.75	0.66	0.65	0.64

3. Selection techniques useful for population improvement

3.1. Utilizing multiple sources of information for selection

Previously, an example of selection for increased NLB resistance in a maize breeding population was given. In this example the selection units and the evaluation units were the same single plants, and only single phenotypic measurments were used as the criteria for selection. This method is called phenotypic mass selection or simply mass selection. Mass selection is the oldest form selection used in breeding. It requires no record keeping and it can be done by simply collecting seeds of desirable plants after pollination. Mass selection done in this manner only selects on females, while the male parent is chosen at random through open pollination. If mass selection is done prior to pollination, selection can be done on both female and male parents by only intermating the selected individuals. Because mass selection is done on single observations only, the selection units are also used as the evaluation units, and generating separate breeding materials for the purpose of evaluation is not needed. The accuracy of mass selection, which is usually low for quantitative traits, is the square root of the narrow sense heritability of the trait of interest in the base population.

In contrast to mass selection where the selection is based on single phenotypic values per selection unit, the criteria for selection may be multiple phenotypic observations combined in a linear fashion into a single value, referred to as a slection index. There are three types of selection indices that were developed so that selection based on the index maximizes genetic

gain. First, the economic selection index was developed by Smith (1936) and Hazel (1943) in which indices are based on phenotypic observations on multiple traits weighted in an optimal way to maximize genetic gain in total net merit. The economic value of each trait, as well as the phenotypic and genetic covariances among traits are incorporated in the weighting procedure. The second is the family selection index developed by Lush (1947) where indices are based on multiple phenotypic observations of a single trait on breeding materials that are related. The index weights in the family selection index take into account the additive genetic covariance among relatives and the heritability of the trait. The third type of selection index combines both information from relatives and information from multiple traits (Henderson, 1963). These selection index weights take into account the economic value of each trait, the phenotypic and genetic covariances among traits, and the additive genetic covariance among relatives.

Multiple phenotypic observations can also be combined using BLUP. BLUP is a procedure that improves upon selection index methods in that it can control for non-genetic factors. In the absence non-genetic factors, BLUP is equivalent to a selection index. With BLUP, breeding materials may be considered unrelated, in which case the BLUP values, referred to as BLUPs, for the breeding materials are based only on their own data. This is referred to as BLUP with genotypes independent and identically distributed (i.i.d). Alternatively, breeding materials may be considered as related based on pedigree or genome-wide marker based relationship (Hayes et al., 2009), referred to as pedigree BLUP and genomic BLUP respectively. Pedigree and genomic BLUP enables the sharing of information among different breeding materials that are related thereby improving selection accuracy.

If multiple traits are phenotyped, multi-trait BLUP, which takes into account the genetic and phenotypic covariances between traits, can be used to utilize all sources of available information. Use of multi-trait BLUP (Henderson, 1973) facilitates the use of economic selection indices because the economic values are directly used to weight the multi-trait BLUPs when estimating net merit for each individual.

Selection indices and BLUPs are estimates of genetic value or they are estimates of breeding value if individuals are assumed related based on additive genetic relationships estimated using pedigree or genome-wide markers. The benefit of using a selection index or BLUP for selection is that that all phenotypic information can be utilized in an optimal way so that selection accuracy is maximized. The accuracy of selection based on a selection index

or BLUP is the square root of the reliability. The reliability depends on the amount and strength of the information utilized as well as heritability of the trait(s).

3.2. Direct and indirect selection

The example maize breeding program described previously illustrated a form of direct selection. With direct selection, the selection criteria (either a phenotypic value, selection index, or BLUP) is based on the trait(s) of interest. For example, if the trait of interest is grain yield in the target environment, selections are made using grain yield data from trials conducted in the target environment. Selection based on traits that may be associated with grain yield such as biomass yield, grain size, number of infloresences, etc. would be considered indirect selection, and such traits are called secondary traits. Likewise, selection based on the trait of interest measured in an environment that does not represent the target environment can be considered indirect selection (Cooper et al., 1993; Cooper and DeLacy, 1994). For instance, if the target environment is tropical, grain yield measured in a temperate environment would be considered a secondary trait and selection based on this trait would be considered indirect selection. Genomic selection is a form of indirect selection because selection is based on a genotype and not directly on phenotypic values.

Indirect selection is often used to eliminate a fraction of the selection units prior to selection on the trait of interest especially when the trait of interest is expensive to measure. For example, in rice breeding, the cost of conducting sensory evaluations to assess grain quality is so high, quality can only be measured on relatively few evaluation units. Therefore, to breed for good or acceptable quality, rice breeders measure and select based on traits which are associated quality such as amylose content.

While indirect selection can be a useful. It has been often used ineffectively. For example, U.S. maize breeding in the early 1900s relied on selection based on appearance of the ear judged at "corn shows" (Bowman and Crossley, 1908) in attempt to develop better yielding maize varieties. In small grain crops, visual selection for yield component traits was often done to try to improve grain yield. This strategy has since been shown to be an ineffective (Rasmusson and Cannell, 1970; McGinnis and Shebeski, 1968).

To determine if indirect selection would be more effective than indirect selection, a metric called the relative efficiency can be computed. The relative efficiency of indirect *vs.* direct selection is simply the ratio of the

expected genetic gain due to indirect selection to the expected gain due to direct selection (Lerner and Cruden, 1947). Assuming the population sizes and number of individuals selected is constant, the relative efficiency of indirect and direct selection depends the selection accuracy of indirect selection as well as the genetic correlation between the indirect and direct selection criteria (**Figure 3**). For the case of indirect selection based on markers, indirect selection accuracy equals one, because a marker genotype is completely heritable. In most cases direct selection is more effective than indirect selection. When much larger population sizes are possible with indirect selection compared to direct selection, there is a greater chance that indirect selection would be a more favorable strategy.

4. Expected genetic gain

Now that genetic gain and the process that generates it has been described, the theory behind why the breeding cycle leads to genetic gain and how rates of genetic gain can be increased will be shown. To do so, two basic genetic principles are explained and then used to show how to predict genetic gain per cycle and per unit time from simple phenotypic mass selection and selection based on selection indices or BLUPs.

4.1. Components of phenotypic value

Characteristics of animals and plants are due to genetic and environmental causes. For example, its well understood that animals or plants that receive better than average nutrition will have better than average growth. This environmental cause of better growth will not be passed down to progeny when these individuals (single animals or plants) are mated. The heritable portion of an individual's observed phenotypic value is its breeding value. Breeding value is the value of an individual as a parent and it is the sum total of the additive effects that an individual's alleles exert on the trait. Only the additive allele effects contribute to breeding value because alleles, and not allelic combinations, are transmitted to progeny due to independent assortment during meiosis.

Although observed phenotypic values are not completely heritable, they can give some indication of breeding value. Using simple linear regression, an individual's breeding value can be predicted using its own phenotypic value for a single trait or from a linear combination of multiple phenotypic values in a selection index or BLUP. The regression formula that allows us to predict

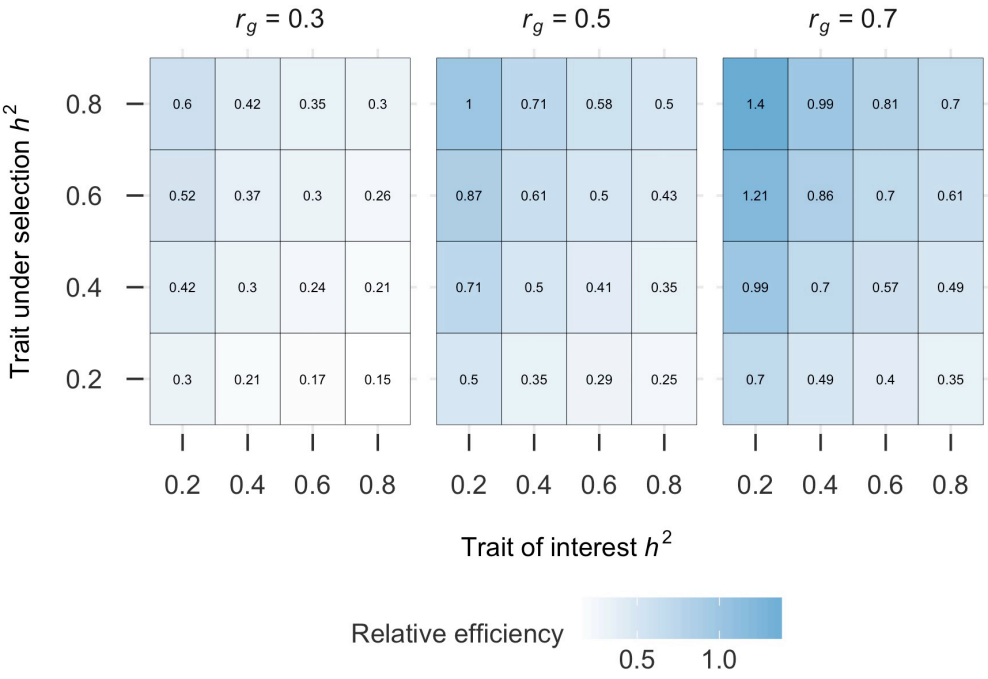

Figure 3: Relative efficiency of indirect *vs.* direct phenotypic selection for different levels of heritability (h^2) of the trait of interest, h^2 of the trait under selection, and additive genetic correlation (r_g) between the two traits. Equal selection intensities are assumed for indirect and direct selection. Indirect selection is more effective than direct selection wherever the relative efficiency values are greater than one.

breeding value based on phenotypic values, selection indices, or BLUPs is $\hat{g}_i = \bar{g} + b_{xg}(x_i - \bar{x})$ where $\hat{g}_i$ is the predicted breeding value of individual i, $\bar{g}$ is the mean breeding value of the population from which i was sampled, b_{xg} is the regression coefficient between breeding value and phenotypic or selection index values, x_i is the phenotypic or selection index value of individual i and $\bar{x}$ is the mean phenotypic or selection index value of the population from which i was sampled.

To express this in a different way, $\bar{g}$ can be subtracted from each side of the equation to get $\hat{g}_i - \bar{g} = b_{xg}(x_i - \bar{x})$. The expression $\hat{g}_i - \bar{g}$ is the genetic superiority of individual i relative to the population mean, and $x_i - \bar{x}$ is the superiority of individual i in terms of its phenotype, selection index, or BLUP, relative to the population mean. In other words, genetic superiority can be predicted from superiority in phenotypic or selection index values, and the degree of correspondence between the two superiorities is the regression coefficient b_{xg}. In the case of simple phenotypic selection, b_{xg} is equal to the the narrow sense heritability, h^2, which is the proportion of the phenotypic variation that can be transmitted to progeny and indicates the effectiveness of selection for realizing genetic gain.

In the same way that genetic superiority of an individual can be predicted, the average genetic superiority for a group of selected individuals can be predicted using $\hat{g}^* - \bar{g} = b_{xg}(x^* - \bar{x})$ where $\hat{g}^*$ is the average breeding value of the selected individuals and x^* is the average phenotypic value of the selected individuals. Later it will be shown how the expression $\hat{g}^* - \bar{g} = b_{xg}(x^* - \bar{x})$ is critical for predicting gain from selection.

4.2. Expected breeding value of progeny and their parents

When two individuals are mated, the average breeding value of their progeny will be equal to the average of their own breeding values. Or in other words, most progeny will be as good as their parents. This is expressed as: $\bar{g}_1 = \frac{1}{2}g_{p_1} + \frac{1}{2}g_{p_2}$, where $\bar{g}_1$ is the average breeding value of a progeny generated from a cross between parent one and two, g_{p_1} is the true breeding value of parent one and g_{p_2} is the true breeding value of parent two. Assuming the parents were not completely inbred, individual progenies will have breeding values that deviate from the average in positive and negative directions (Hill, 1993; Franklin, 1977) because progeny do not receive the same set of alleles from each parent.

4.3. The Breeder's equation

The breeder's equation allows one to predict gain from selection and it can be derived using the two basic principles just described. Assume that unbiased estimates of breeding value are available, then estimated breeding values of parents can be substituted for true breeding values in $\bar{g}_1 = \frac{1}{2}g_{p_1} + \frac{1}{2}g_{p_2}$ giving $\bar{g}_1 = \frac{1}{2}\hat{g}_{p_1} + \frac{1}{2}\hat{g}_{p_2}$. Thus, its possible to predict the population mean breeding value of a population of progeny simply based on estimated breeding values of the parents. In other words, $\hat{g}^*$, which was defined earlier as the average estimated breeding value for a group of selected individuals, would be equal to the average breeding value of the progeny of these selected individuals, denoted as $\bar{g}_1$, if they were mated. Substituting $\bar{g}$ for $\hat{g}^*$ in the prediction equation mentioned above we can arrive at $\bar{g}_1 - \bar{g} = b_{xg}(x^* - \bar{x})$. Notice that $\bar{g}_1 - \bar{g}$ is simply the difference in the mean breeding value that occurred due to one cycle of selection. This is also referred to as the response, R, to one cycle of selection. Substituting R for $\bar{g}_1 - \bar{g}$, response to one cycle of selection can be predicted using $R = b_{xg}(x^* - \bar{x})$. In the case of phenotypic mass selection, recall that $b_{xg} = h^2$, and the formula for one response to one cycle of selection can be expressed as $R = h^2(x^* - \bar{x})$. The expression $x^* - \bar{x}$ is commonly referred to as the selection differential, S. Substituting S for $(x^* - \bar{x})$ we arrive at the $R = h^2S$, which is a commonly used formula for the prediction of response to phenotypic mass selection. When selection is imposed differently on males and females. Response is predicted by $R = \frac{1}{2}R_m + \frac{1}{2}R_f$. Where R_m is the expected response in males and R_f is the expected response in females. Phenotypic mass selection after flowering imposes no selection in males, thus the expected response in this case is $R = \frac{1}{2}h^2S$

4.4. Equivalent versions of the Breeder's Equation

Other algebraic rearrangements of $\hat{g}_i - \bar{g} = b_{xg}(x_i - \bar{x})$ and $\bar{g}_1 = \frac{1}{2}g_{p_1} + \frac{1}{2}g_{p_2}$ allow us to express the breeders equation in different ways (**Table 2**) which are useful for for predicting genetic gain per cycle in different scenarios. To demonstrate, lets assume that selection is based on estimated breeding values obtained using BLUP or a selection index. Also assume that the selection, evaluation, and recombination units are the same individuals and selection is imposed equally on both males and females.

Recall that the average breeding value for a group of selected individuals can be predicted using $\hat{g}^* = \bar{g} + b_{xg}(x^* - \bar{x})$, where $\hat{g}^*$ is the average breeding value, $\bar{g}$ is the average breeding value of the population, b_{xg} is the regression

coefficient between breeding values and estimated breeding values, and x^* is the average estimated breeding value for the selected individuals, and $\bar{x}$ is the average estimated breeding value of the population.

Also recall that $x^* - \bar{x}$ is also known as the selection differential, S. Thus the prediction equation can be written as $\hat{g}^* = \bar{g} + b_{xg}S$. The selection differential, S, can be expressed in units of standard deviation, referred to as k, using $k = S/\sigma_x$ where σ_x is the additive genetic standard deviation of the estimated breeding values. Through rearrangement, $S = k\sigma_x$. Substituting $k\sigma_x$ for S in $\hat{g}^* = \bar{g} + b_{xg}S$ we arrive at $\hat{g}^* = \bar{g} + b_{xg}k\sigma_x$. Subtraction of $\bar{g}$ from both sides gives $\hat{g}^* - \bar{g} = b_{xg}k\sigma_x$. Recall that $\hat{g}^*$ is equal to $\bar{g}_1$, therefore $\hat{g}^* - \bar{g}$ becomes $\bar{g}_1 - \bar{g}$ which is simply the change in population means due to selection, R. Through substitution of R for $\hat{g}^* - \bar{g}$ in $\hat{g}^* - \bar{g} = b_{xg}k\sigma_x$, we arrive at $R = b_{xg}k\sigma_x$

It is more useful to express the breeder's equation in terms of the selection accuracy, r_{xg}, which is the correlation between the selection indices or BLUPs and the breeding values. To do so, we start with the definitions of regression and correlation coefficients in terms of variance and covariance: $r_{xg} = \sigma_{xg}/\sigma_g\sigma_x$ and $b_{xg} = \sigma_{xg}/\sigma_x^2$. The parameter σ_x^2 can be expanded to $\sigma_x\sigma_x$, thus $b_{xg} = \sigma_{xg}/\sigma_x\sigma_x$. Then based on the definition of r_{xg}, we can substitute σ_{xg}/σ_x with $r_{xg}\sigma_g$ giving $b_{xg} = r_{xg}\sigma_g/\sigma_x$. Now, $r_{xg}\sigma_g/\sigma_x$ can be substitued for b_{xg} in $R = b_{xg}k\sigma_x$ giving $R = k\sigma_x r_{xg}\sigma_g/\sigma_x$. Then σ_x cancels out giving: $R = r_{xg}\sigma_g k$. If the parameters in $R = r_{xg}\sigma_g k$ can be assumed to remain constant over cycles of selection then R is the expected response, also called expected genetic gain from selection per cycle. The expected gain per unit time, R_t, referred to as the rate of genetic gain, is simply R/L, where L is the length of time required to complete one breeding cycle.

Note that several of these equations refer to the selection intensity k which is the selection differential is expressed in units of standard deviation. This is useful because assuming the trait phenotypes are normally distributed, k can be estimated based on percentage of individuals selected as parents from the population of selection units (referred to as percent selected) using properties of the normal distribution. The relationship between the percent selected and the selection intensity is shown in **Figure 4**. In breeding programs, the percent selected, and consequently the selection intensity, are often increased by increasing the number of selection units that are evaluated.

The breeder's equations just described can only accurately predict genetic gain in quantitative traits per cycle or per unit time assuming that the selection units, evaluation units, and recombination units are the same breeding

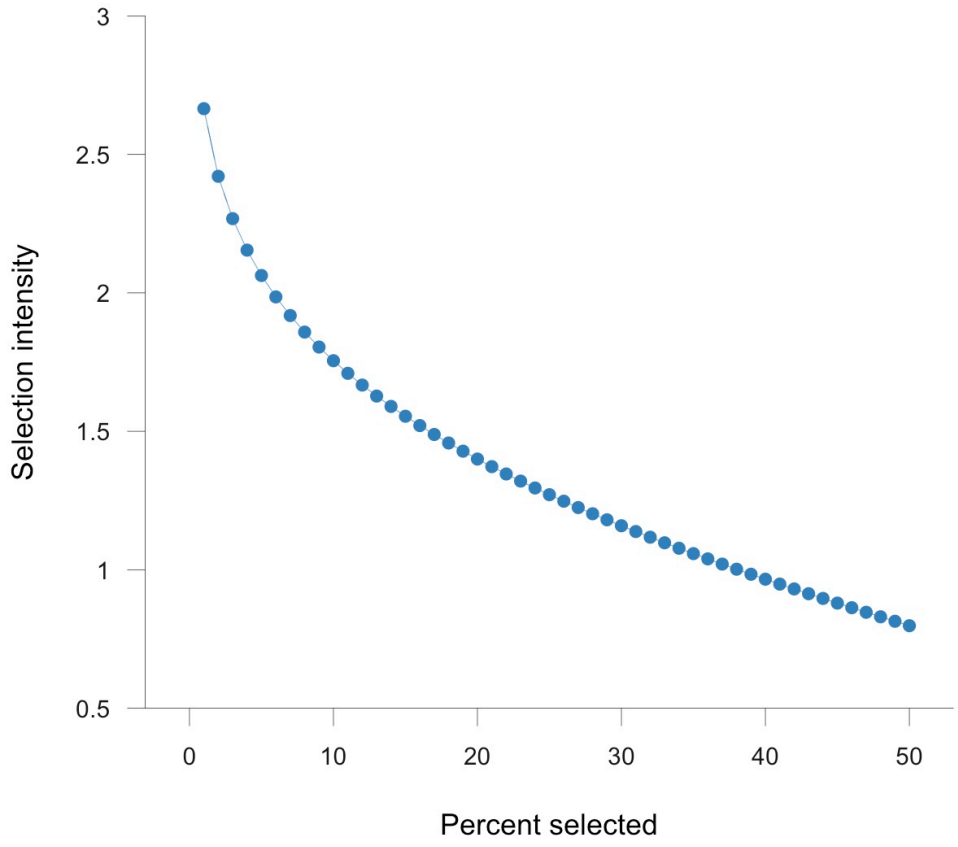

Figure 4: Selection intensity *vs.* percent selected for percent values ranging from 1 to 50

Table 2: Useful formulas for predicting genetic gain

Equation	Description
$R = h^2 S$	Expected gain from phenotypic mass selection when the selection differential, S, is known
$R = kh^2\sigma_p$	Expected gain from phenotypic selection
$R = kr_{xg}\sigma_g$	Expected gain per cycle based on multiple observations combined using a selection index or BLUP
$R = kh_i h_j r_g \sigma_{p_i}$	Expected gain per cycle in trait i based selection based phenotypic values of trait j
$R = kh_j r_g \sigma_{g_i}$	Expected gain per cycle in trait i based on multiple observations of trait j combined using a selection index or BLUP j
$R = \frac{1}{2}R_m + \frac{1}{2}R_f$	Overall expected gain when selection is imposed differently on males and females
$R_t = R/L$	Expected gain per unit time based on expected gain per cycle and the breeding cycle duration L

h^2 is the heritability, σ_p is the phenotypic standard deviation, k is the selection intensity, r_{xg} is correlation between true and estimated breeding values referred to as the selection accuracy, σ_g is the additive genetic standard deviation, h is the square root of the heritability, r_g is the additive genetic correlation between traits i and j, R_m and R_f is the genetic gain per cycle in males and females respectively.

materials, selection is done in one stage, generations do not overlap, and the breeding population is reasonably large (Burrows, 1972). Accurate prediction of genetic gain when these conditions are violated, which is often the case in breeding programs, requires additional formulas. However, the breeder's equations are still useful for understanding the factors that affect genetic gain and their relative importance. For example, it is clear that the accuracy of selection is an important factor for achieving genetic gain. As shown in **Figure 5**, there is a direct linear increase between the selection accuracy and the rate of genetic gain per cycle for all levels of percent selected. Decreasing levels of the percent selected are also associated with greater rates of genetic gain per cycle, as shown in **Figure 6**, but the relationship is not linear and it is weaker when the selection accuracy is low. Another critical variable is

the the duration of the breeding cycle which has an indirect relationship with the rate of genetic gain per unit time. For example, reducing the breeding cycle from six years to three years would double the rate of genetic gain per unit time, assuming the accuracy, additive genetic variance, and percent selected, remain unchanged. Of all the variables in the breeder's equation, the length of the breeding cycle duration has probably the greatest potential to be manipulated to increase rates of genetic gain in plant breeding programs.

4.5. Stability of genetic variance

Predictions of per-cycle gain from selection also assume that parameters in the breeder's equation remain constant over cycles. Genetic variance remains constant over cycles assuming that the trait under selection conforms follows the infinitesimal model of inheritance (Crow and Kimura, 1970), which states that there are an infinite number of loci, each with a small, additive effect on the trait (Fisher, 1918). Selection does in fact affect the genetic variance, a phenomenon referred to as the Bulmer effect (Bulmer, 1971). Immediately after selection and prior to recombination, genetic variance is reduced depending on the selection accuracy and the percent of the population selected. **Figure 7** shows the proportion of the additive genetic standard deviation remaining after selection for different values of selection accuracy and percent selected. However, much of the genetic variance is restored during recombination. After a few cycles of selection a balance between loss of variance induced by selection and generation of variance due to recombination is obtained (Bulmer, 1971; Gomez-Raya and Burnside, 1990), thus it is reasonable to assume that the genetic variance will remain constant for a reasonable length of time if the trait under selection is quantitative and the population has already been under selection for a few cycles.

5. Realized genetic gain

5.1. Examples of genetic gain realized

Hundreds of selection experiments have been conducted since the early 1900s. More than 60 selection experiments conducted using animal species have been summarised in a 1988 review by Sheridan (1988), and more than 50 selection experiments conducted in self and cross-pollinated crop species have been reviewed by Hallauer and Darrah (1985). Outcomes of selection experiments have confirmed that the rate of genetic gain per cycle depends on the selection accuracy, selection intensity, and additive genetic variance

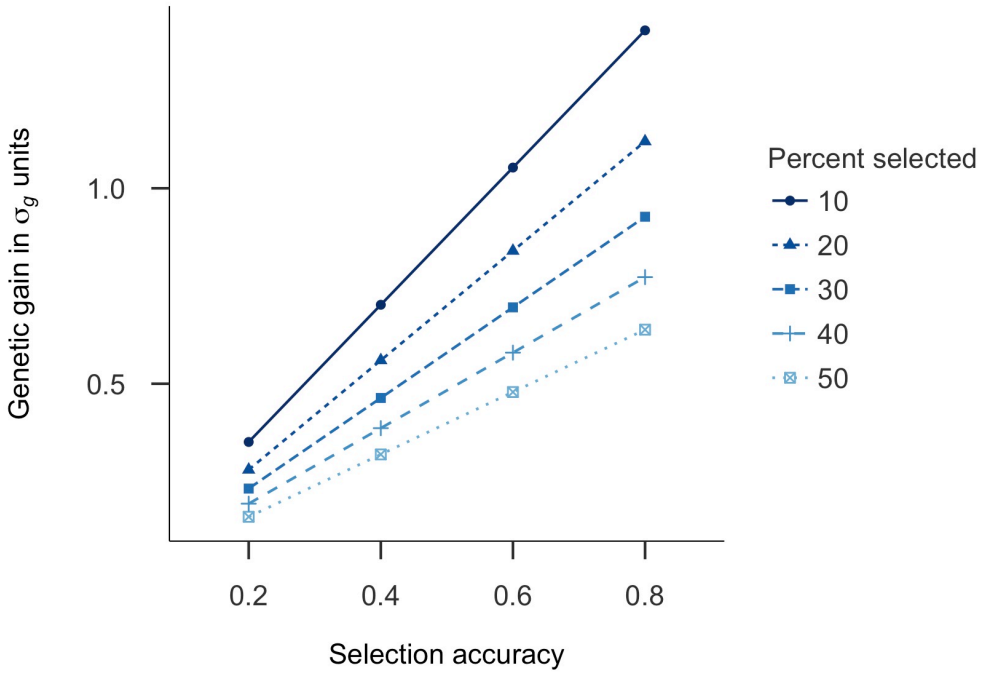

Figure 5: The rate of genetic gain per cycle with increasing levels of selection accuracy. A clear linear relationship between the selection accuracy and the rate of genetic gain in units of genetic standard deviation is demonstrated.

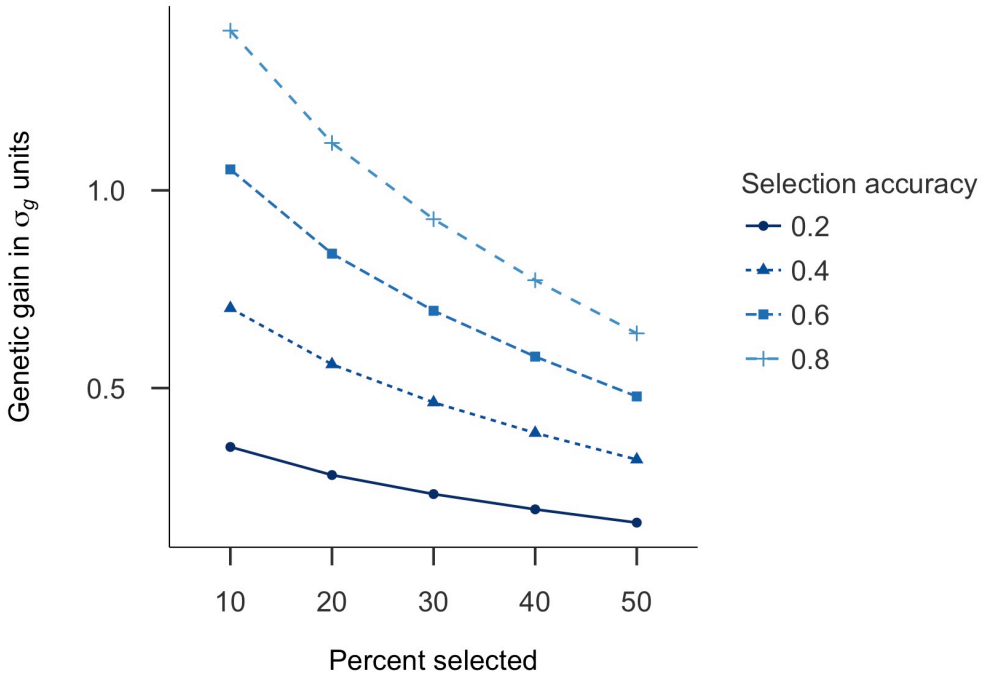

Figure 6: The rate of genetic gain per cycle with decreasing levels of percent selected. Lower values of percent selected are associated with greater rates of genetic gain per cycle, especially when the selection accuracy is moderate to high.

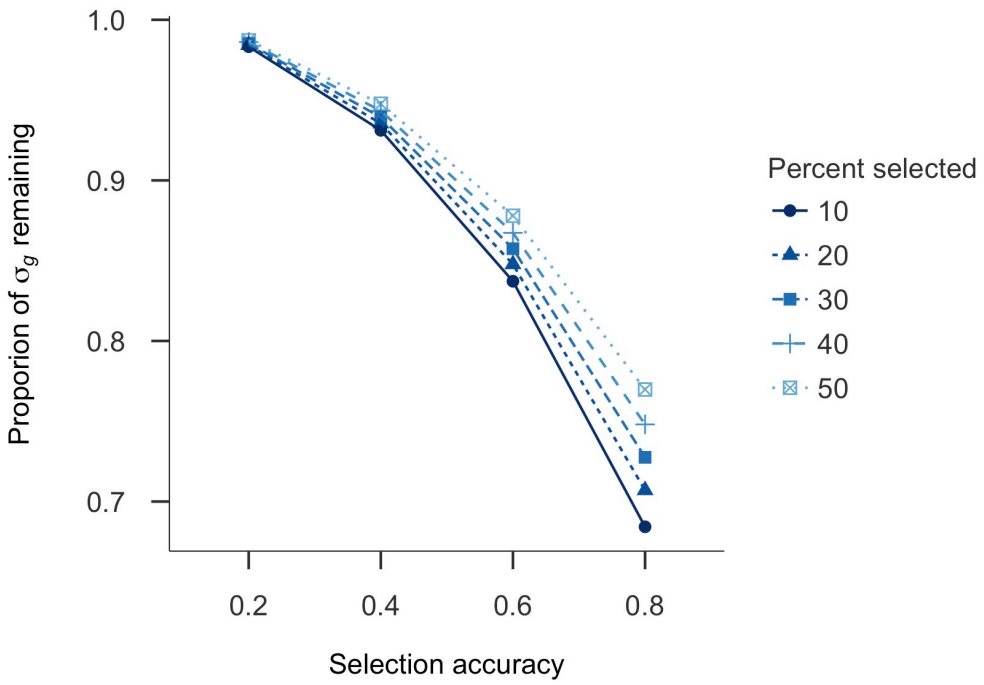

Figure 7: The proportion of the additive genetic standard deviation remaining immediately after truncation selection and prior to recombination for different values of selection accuracy and percent selected. As selection accuracy increases the additive genetic standard deviation remaining after selection decreases.

as expected based on theory. Random fluctuations in selection responses are also observed, confirming theoretical expectations of variability in selection response due to genetic drift (Wright, 1930).

In addition to confirming theoretical expectations, selection experiments have also demonstrated the impressive ability of populations to respond to selection over the long term. For example, in the Illinois Long-Term Selection experiment in maize (*Zea mays* L.), recurrent selection for high percent oil and high percent protein in maize kernels has produced consistent changes in phenotype for more than 100 cycles of selection. For example, at the start of the Illinois Long-Term Selection experiment experiment in 1896 average percent oil was 4.69, and by 2004 the population being selected for higher percent oil reached an average of 22.1 percent oil (University of Illinois Urbana-Champaign Department of Crop Science, 2007a,b). In addition to the changes in means, the ranges of the populations observed in 1896 and 2004 are also drastically different (**Figure 8**). This demonstrates the effectiveness of selection within a single population for bringing about major changes in the phenotypic means and ranges.

The long-term genetic gain from selection observed in the Illinois Long-Term Selection experiment is consistent with a hypothesis that there are many loci affecting oil and protein content, and alleles that have a positive effect on these traits were relatively rare in the initial population (Dudley, 2007). Similar observations have been observed in animal species. For example, in the Virginia Chicken lines selection experiment, continued response to selection for body weight was observed for 38 cycles, and response to selection in this experiment was found to be due the effect of many small effect loci (Jacobsson et al., 2005).

Genetic gain has not only been observed in controlled experiments. Several applied animal and crop breeding programs have also documented gain from selection across many years of selective breeding (Duvick, 2005; Piepho et al., 2014; Laidig et al., 2014; Burnside and Legates, 1967; Chen et al., 2003; Cloete et al., 2004). In summary, nearly 100 years of experimentation and breeding has confirmed that if a trait is heritable it will respond to selection and it will continue to respond to selection for many cycles of selection if conferred by many loci.

5.2. Estimating realized genetic gain

The success of a plant breeding program has traditionally been assessed based on the adoption of the varieties it has released (Brennan and Byerlee,

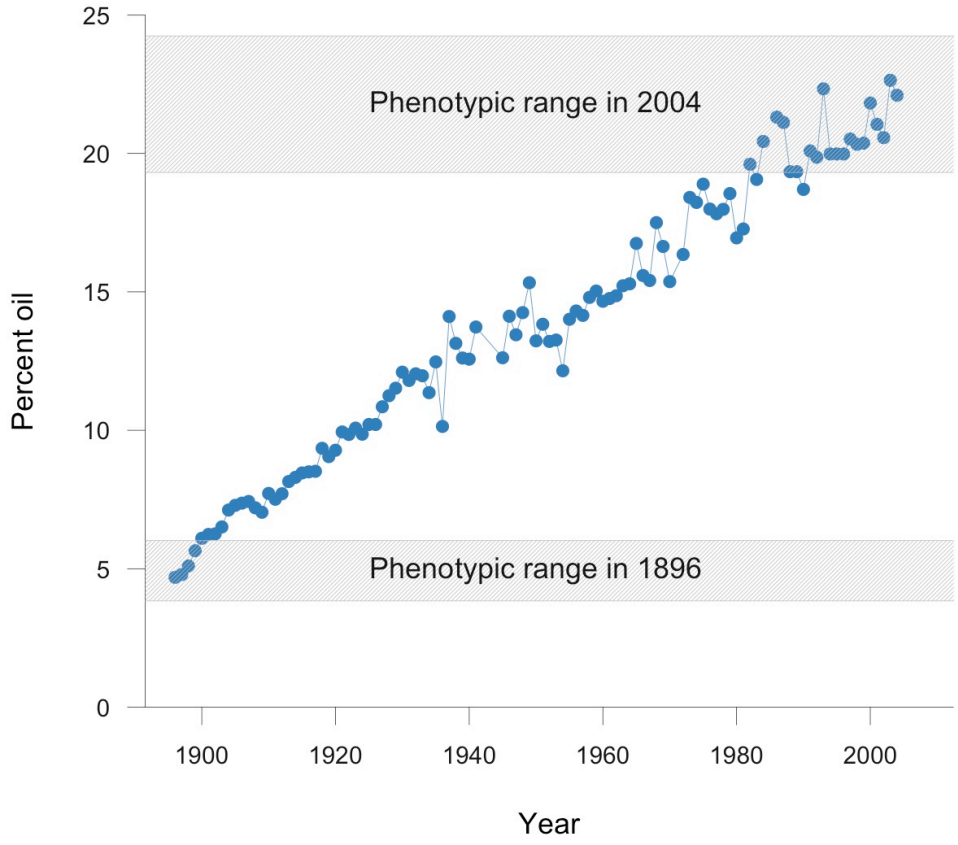

Figure 8: Realized genetic gain from selection for increased percent oil of kernels in maize in the Illinois Long-Term Selection experiment.

1991) because its ultimate goal is to develop varieties grown by farmers; however, variety adoption is also heavily affected by the efficiency of the seed system and can only be assessed years after variety release. The genetic trend, or realized rate of genetic gain from a breeding program could be a useful indicator of a breeding program's success independent of the seed system. On the other hand, realized rates of genetic gain are affected by factors that are not under one's control including the genetic nature of the target traits, and genetic drift in addition to those under the breeder or breeding organization's control including the design of the breeding program and its operational efficiency. Estimates of realized genetic gain can also only be assessed after many years of breeding. Thus, although the rate of genetic gain realized may be a good indicator of a breeding program's success, it may be a poor indicator of the performance of the breeder or breeding organization.

Measuring the rate of genetic gain realized by a breeding program for individual traits or for net merit is conceptually straightforward. All that is required are estimates of the mean breeding value for the trait(s) of interest for the breeding population per year or per breeding cycle. The estimate of realized gain per year or per cycle, is simply the slope of the regression line of mean breeding value on year or cycle number (Garrick, 2010; Eberhart, 1964).

In practice, accurately estimating the rate of genetic gain realized by a breeding program can be problematic because non-genetic trends due to changes in agronomic practices or climate change are confounded with genetic trends. For example, increases in minimum night temperature are known to negatively impact yield in both rice (*Oryza sativa*) and wheat (*Triticum aestivum* L.). For these crops, change in average yield observed over time may be the result of a reduction in average yield due to climate change and an increase in average yield due to breeding.

To overcome the problem of confounding, breeding materials from different cycles or years can be phenotypically evaluated in a common set of environments. Then the trend in the average phenotypic value over breeding cycles or years must be due change in average breeding value over time, assuming the breeding materials were developed through the population improvement process. Ideally, the phenotypic evaluation is done using recently multiplied seed, to avoid possible confounding of genetic trend with seed age, and using large samples of breeding materials from each year or cycle of breeding to minimize sampling error (Hallauer et al., 2010). This is often done

when estimating realized genetic gain from selection experiments. Unfortunately, most applied breeding programs keep only the most elite breeding materials for a short period of time.

In a series of experiments conducted conducted between 1977 and 2004 to assess genetic contributions to improvement in U.S. corn yields, Duvick (2005) evaluated varieties released over many years in the same environments and regressed variety mean yield on the year of release. This method, referred to as the "ERA trial" method, has become popular for estimating rates of genetic gain realized by plant breeding programs because seed of varieties released over time is more readily available than large samples of breeding materials representing multiple cycles or years of breeding. Many examples of this approach are available (Peng et al., 2000; Donmez et al., 2001; Brancourt-Hulmel et al., 2003). However, the ERA trial method does not necessarily provide an accurate estimate of the rate of genetic gain because released varieties are not representative of the breeding population.

Historical breeding trial data could be useful for estimating realized genetic gain for the trait(s) of interest, like what is done in livestock (Garrick, 2010). In fact, some studies have used crop variety trial data to estimate realized gain (Piepho et al., 2014; Laidig et al., 2014; Mackay et al., 2011). However, historical plant breeding trial datasets often do not meet the requirements necessary to obtain a reliable estimates of mean genetic value per year or cycle. The most frequent problem is confounding of genetic and year effects due to a lack of genetic connectivity between years caused by testing only a small proportion of breeding materials for more than one or two years. Estimates of mean breeding value per year estimated using pedigree or genomic relationship may be less confounded with year because genetic connectivity between years can be increased through the pedigree or genomic relationship. If a control population or multiple check varieties were grown in all years, then data on the the control population can be used to help separate the genetic and non-genetic effects over time. However, it is relatively rare to find an applied breeding program that has evaluated a control population or the same set of checks for several consecutive years.

6. How to improve rates of genetic gain

Although theory indicates that realizing genetic gain is possible as long as a population improvement strategy is followed and the traits of interest are heritable, there are also a series of components that should be in place

in a breeding program to ensure that genetic gain can be realized. Once these components are well established, a breeding program can then incorporate additional components to help it achieve a higher rate of genetic gain. Following the example of Maslow's hierarchy of needs (Maslow, 1943), what is needed in a breeding program to realize genetic gain and then improve it further can be described as having a hierarchical structure (**Figure 9**) with different components building on each other to achieve a more effective breeding program.

6.1. Foundational components

If a breeding program follows a generate-evaluate-select-recombine process **Figure 1**, it can expect to realize genetic gain as long as a set of essential components are in place. Three of these essential components that are the foundation for all others are 1) clear objectives, 2) meticulous data management, and 3) effective operations. Once these are in place, it is possible to achieve the next level of required components, 4) effective phenotyping and 5) selection based on BLUP.

Having clear objectives enables the direction of the breeding program to set towards a specific goal. A breeding program without clear objectives is like a ship travelling without a compass. No matter how fast or efficiently the ship moves, it will not reach any particular destination. When setting clear objectives, the target traits and their relative importance should be well defined. Defining the set of environments experienced by the farmers that the breeding program aims to serve, referred to as the target population of environments (TPE) (Cooper et al., 1993; Cooper and DeLacy, 1994), is also required. Because genetic gain is realized in small increments over time, it is important that the objectives of the breeding program not change frequently, otherwise a meaningful level of improvement in the target traits in the TPE will not be achieved. Determining the economic weights for each target trait helps to remove ambiguity in the breeding objectives and reduces human error in the process of selection towards these objectives.

Meticulous data management is needed to ensure that the breeding program's phenotypic, genotypic, and pedigree data can be used for selection and to avoid losses in selection accuracy due to errors in data management. Without this component, resources used to generate and utilize data will be wasted. For example, if phenotypic data is kept in simple electronic spreadsheets, a small mistake in sorting data or combining different sources of information could lead to zero selection accuracy. As another example, if

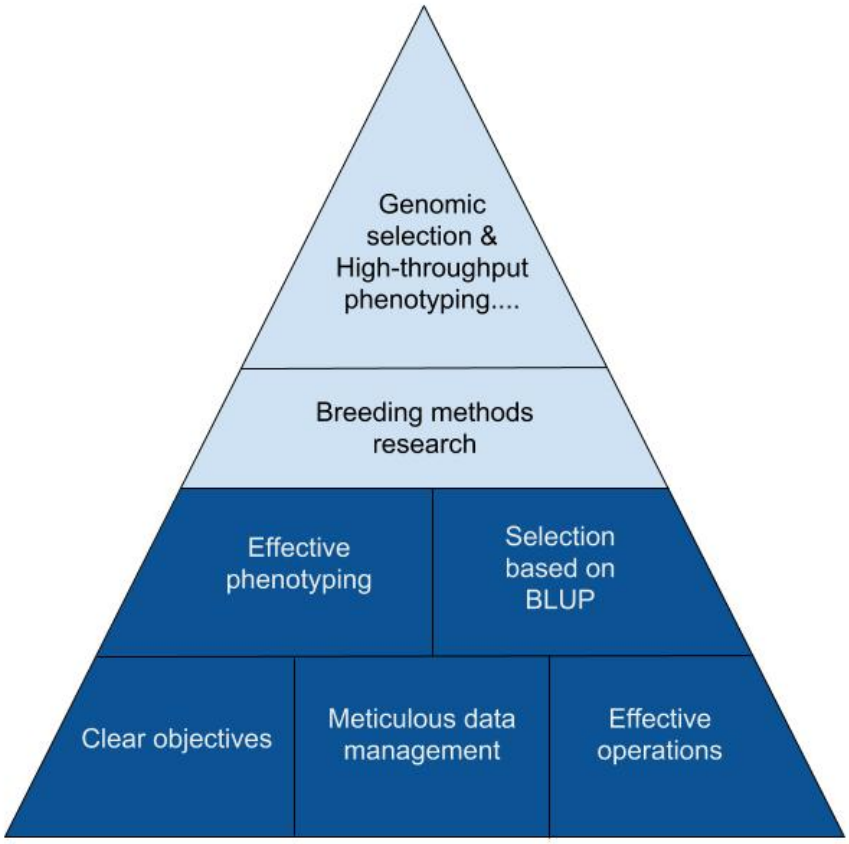

Figure 9: A Breeding program's hierarchy of needs. Each section of the pyramid represents a different component that is important for realizing genetic gain. The first two levels, shaded in a darker color, consists of the foundational components that must be in place to ensure that genetic gain can be realized. The upper part of the pyramid consists of advanced components that can help increase the rate of genetic gain realized by the breeding program.

spreadsheets containing plot yield data do not also include harvested area of the plots, the yield data may no longer be usable. Proper use of a breeding data management system can help in avoiding data such errors. Several such data management systems are now available.

Effective operations means that routine activities such as planting, harvesting, crossing, and managing germplasm are done correctly so that the identity of the breeding materials is preserved during the breeding process. If a breeding program cannot execute routine activities correctly, it will loose control over which breeding materials ultimately get selected and recombined. Mistakes during the execution of routine activities can be minimized by ensuring that population sizes are manageable given the resources available, by ensuring staff are adequately trained and managed, and by consistently following a set of well thought out standard operating procedures. Detecting mistakes when they occur is also important so that they can be mitigated and prevented in the future. Morphological checks, visually distinct plants or families that can be easily identified, can help detect errors in planting and harvesting. If these checks are not planted in the correct plots or if they are mislabeled after harvest, this can indicate that a mistake occurred during the field operations and the data and germplasm resulting from these operations may be unusable. Genotyping can also be useful for detecting seed mixtures or failed crosses so that they may be discarded (Chen et al., 2016).

Effective phenotytpic evaluations are needed to ensure that the breeding program will realize genetic gain in the target traits according to the objectives of the breeding program. For phenotytpic evaluations to be effective they should be done in a way so that the phenotypic data collected is predictive of that which would be observed in farmers' fields. This requires proper phenotyping methods, and appropriate statistical design which includes conducting trials in multiple environments that adequately sample the TPE. For example, if the target trait is grain yield and farmers grow the crop in their fields as a uniform stand (as opposed to a mixture), then grain yield should be measured in field conditions, using field sites that that represent the farmer's environments, and using homogenous breeding materials grown in reasonably large plots (Rebetzke et al., 2014). Research to assess patterns of genotype by environment interaction (GxE) and different phenotyping methods can help in determining what field trial sites to use and how to conduct phenotypic evaluations for improving the traits of interest in the TPE. In many crops, research studies on phenotyping methods for various traits have already been conducted and can be readily adopted.

While it is possible to use single phenotypic observations for selection, selection based on BLUP is critical for improving low heritability traits and traits that are affected by GxE such as yield. Assuming that the statistical design used for phenotypic evaluation is adequate, phenotypic data across multiple years and environments can be combined using BLUP to improve selection accuracy. Selection accuracy can also be improved by using pedigree and/or marker information in BLUP. If the breeding program targets multiple traits, selection based on multi-trait BLUPs combined using economic weights is optimal (Hazel, 1943). Finally, it is critical that BLUPs or an index of multi-trait BLUPs be the criteria for selection such that the average BLUP or index of the selected breeding materials is superior to that of the population average.

6.2. Advanced components

Once a population improvement-based breeding program has all the required components in place to ensure that it can realize genetic gain, it can begin to conduct reasearch aimed at improving the breeding process. For example, empirical and/or simulation experiments can be used to evaluate techniques for reducing the time needed to generate breeding materials, such as off-season nurseries, greenhouses or doubled haploids; and to evaluate different ways of leveraging techniques such as genomic selection and high-throughput phenotyping in the breeding cycle. Before making changes to the breeding process, this kind of "breeding methods research" is critical to ensure that the proposed changes will actually improve the breeding program's efficiency or rate of realized genetic gain. Examples of recent breeding methods research studies include Heffner et al.; Endelman et al. (2014); Lorenz (2013); Rutkoski et al. (2015); Massman et al. (2013); Longin et al. (2015)

Simulation is useful tool which allows testing different hypotheses about breeding methods quickly and at low cost. Simulations many be deterministic, relying on formulas for predicting genetic gain such as the breeders equation, or they may be stochastic, relying computers to model each process in the breeding cycle. Stochastic simulations are useful when the breeding scenarios to be tested are complex. Several stochastic simulation platforms useful for plant breeding are now available such as Breeding Scheme Language (Yabe et al., 2017), AphaSim (Faux et al., 2016) and QuGene (Podlich and Cooper, 1998).

7. Conclusion

The objective of this chapter was to increase awareness and knowledge about genetic gain so that breeding programs can begin to take actions that will enable them to be more successful. Genetic gain was defined as the improvement in average genetic value in a population or the improvement in average phenotypic value due to selection within a population over cycles of breeding assuming that the effect of environment remains constant. The importance achieving genetic gain over cycles of selection was demonstrated using a simple simulation, where achieving genetic gain over multiple cycles of selection was shown to dramatically increase the probability of identifying promising new varieties on an ongoing basis compared to repeated selecting from a population that is not improving such as a germplasm collection. Next, key selection techniques such as combining multiple sources of information for selection using selection indices and BLUP and direct and indirect selection were explained and discussed to help point out when these techniques are useful. Then, to show that genetic gain is readily achievable, quantitative genetic principles were used to explain how genetic gain can be predicted from one generation to the next. This was followed by a discussion of realized genetic gain including impressive examples from selection experiments, and a review of methods used to estimate realized rates of genetic gain. Lastly, a hierarchy of different components needed in a breeding program was provided, first to ensure that the breeding program can realize genetic gain and second to enable the breeding program to realize genetic gain at a faster rate.

8. Acknowledgements

This work was supported by the Bill and Melinda Gates Foundation project: Transforming Rice Breeding and Excellence in Breeding Platform (EiB) of the Consultative Group on International Agricultural Research (CGIAR).

References

Allaire, F.R., 1980. Mate selection by selection index theory. Theoretical and Applied Genetics 57, 267–272. doi:10.1007/BF00264953.

Bernardo, R., 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Science 48, 1649–1664.

769 Bowman, M.L., Crossley, B.W., 1908. Corn. Kenyon, Des Moines, IA.

770 Brancourt-Hulmel, M., Doussinault, G., Lecomte, C., Bérard, P., Le Buanec,
771 B., Trottet, M., 2003. Genetic improvement of agronomic traits of win-
772 ter wheat cultivars released in France from 1946 to 1992. Crop Sci-
773 ence 43, 37–45. URL: http://dx.doi.org/10.2135/cropsci2003.3700,
774 doi:10.2135/cropsci2003.3700.

775 Brennan, J., Byerlee, D., 1991. The rate of crop varietal replacement on
776 farms: measures and empirical results for wheat 4, 99–106.

777 Bulmer, M.G., 1971. The effect of selection on genetic vari-
778 ability. The American Naturalist 105, 201–211. URL:
779 http://www.jstor.org/stable/2459550, doi:10.2307/2459550.

780 Burnside, E., Legates, J., 1967. Estimation of genetic trends in dairy cattle.
781 Journal of Dairy Science 50, 1448–1457.

782 Burrows, P.M., 1972. Expected welection eifferentials for
783 directional selection. Biometrics 28, 1091–1100. URL:
784 http://www.jstor.org/stable/2528642, doi:10.2307/2528642.

785 Chen, J., Zavala, C., Ortega, N., Petroli, C., Franco, J., Burgueño, J., Cos-
786 tich, D.E., Hearne, S.J., 2016. The Development of quality control geno-
787 typing approaches: A Case study using elite maize lines. PLOS ONE 11,
788 e0157236. URL: https://doi.org/10.1371/journal.pone.0157236.

789 Chen, P., Baas, T.J., Mabry, J.W., Koehler, K.J., Dekkers, J.C., 2003. Ge-
790 netic parameters and trends for litter traits in US Yorkshire, Duroc, Hamp-
791 shire, and Landrace pigs. Journal of Animal Science 81, 46–53.

792 Cloete, S.W.P., Gilmour, A.R., Olivier, J.J., Van Wyk, J.B., 2004. Ge-
793 netic and phenotypic trends and parameters in reproduction, greasy fleece
794 weight and liveweight in Merino lines divergently selected for multiple rear-
795 ing ability. Australian Journal of Experimental Agriculture 44, 745–754.

796 Collard, B.C.Y., Beredo, J.C., Lenaerts, B., Mendoza, R., Santelices,
797 R., Lopena, V., Verdeprado, H., Raghavan, C., Gregorio, G.B.,
798 Vial, L., Demont, M., Biswas, P.S., Iftekharuddaula, K.M., Rah-
799 man, M.A., Cobb, J.N., Islam, M.R., 2017. Revisiting rice breeding
800 methods evaluating the use of rapid generation advance (RGA)

for routine rice breeding. Plant Production Science 1008, 1–16. URL: https://www.tandfonline.com/doi/full/10.1080/1343943X.2017.1391705, doi:10.1080/1343943X.2017.1391705.

Cooper, M., Byth, D.E., DeLacy, I.H., Woodruff, D.R., 1993. Predicting grain yield in Australian environments using data from CIMMYT international wheat performance trials. 1. Potential for exploiting correlated response to selection. Field Crops Research 32, 305–322. doi:10.1016/0378-4290(93)90039-P.

Cooper, M., DeLacy, I.H., 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theoretical and Applied Genetics 88, 561–572. doi:10.1007/BF01240919.

Crow, J., Kimura, M., 1970. An introduction to population genetics theory. Harper & Row, New York, NY.

Cullis, B.R., Smith, A.B., Coombes, N.E., 2006. On the design of early generation cultivar trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics 11, 381–393.

Donmez, E., Sears, R.G., Shroyer, J.P., Paulsen, G.M., 2001. Genetic gain in yield attributes of winter wheat in the great plains. Crop Science 41, 1412–1419. URL: http://dx.doi.org/10.2135/cropsci2001.4151412x, doi:10.2135/cropsci2001.4151412x.

Dudley, J.W., 2007. From means to QTL: The Illinois long-term selection experiment as a case study in quantitative genetics. Crop Science 47. doi:10.2135/cropsci2007.04.0003IPBS.

Duvick, D.N., 2005. The contribution of breeding to yield advancements in maize (Zea mays L.). Advances in Agronomy 86, 83–145.

Eberhart, S.A., 1964. Least squares method for comparing progress among recurrent selection methods. Crop Science 4, 230–231.

Endelman, J.B., Atlin, G.N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M.E., Jannink, J.L., 2014. Optimal design of preliminary yield trials with genome-wide markers. Crop Science 54, 48–59. doi:10.2135/cropsci2013.03.0154.

Faux, A., Gorjanc, G., Gaynor, R., Battagin, M., Edwards, S., Wilson, D., Hearne, S., Gonen, S., Hickey, J., 2016. AlphaSim: software for breeding program simulation. The Plant Genome 9.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Philosophical Transactions of the Royal Society of Edinburgh 52, 339–433.

Fisher, R.A., 1930. The Fundamental theorem of natural selection, in: The Genetical theory of natural selection. The Clarendon Press, Oxford. chapter II, pp. 22–47.

Franklin, I.R., 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. Theoretical Population Biology 11, 60–80. URL: http://www.sciencedirect.com/science/article/pii/0040580977900077, doi:https://doi.org/10.1016/0040-5809(77)90007-7.

Gallais, A., Bordes, J., 2007. The Use of doubled haploids in recurrent selection and hybrid development in maize. Crop Science 47, S–190–S–201. URL: http://dx.doi.org/10.2135/cropsci2007.04.0019IPBS, doi:10.2135/cropsci2007.04.0019IPBS.

Garrick, D.J., 2010. An animal breeding approach to the estimation of genetic and environmental trends from field populations. Journal of animal science 88. doi:10.2527/jas.2009-2329.

Gianola, D., 2013. Priors in whole-genome regression: The Bayesian alphabet returns. Genetics 194, 573–596. doi:10.1534/genetics.113.151753.

Gomez-Raya, L., Burnside, E.B., 1990. The effect of repeated cycles of selection on genetic variance, heritability, and response. Theoretical and Applied Genetics 79, 568–574. URL: https://doi.org/10.1007/BF00226169, doi:10.1007/BF00226169.

Goulas, C.K., Lonnquist, J.H., 1976. Combined half-sib and S1 family selection in a maize composite population. Crop Science 16, 461–464. URL: http://dx.doi.org/10.2135/cropsci1976.0011183X001600040005x, doi:10.2135/cropsci1976.0011183X001600040005x.

Haghighattalab, A., Pérez, L.G., Mondal, S., Singh, D., Schinstock, D., Rutkoski, J., Ortiz-Monasterio, I., Singh, R.P., Goodin, D., Poland, J., 2016. Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods 12, 35.

Hallauer, A.R., Carena, M.J., Miranda Filho, J.B., 2010. Quantitative Genetics in Maize Breeding. 3rd ed., Springer-Verlag New York, LLC.

Hallauer, A.R., Darrah, L.L., 1985. Critical Reviews in Plant Sciences Compendium of recurrent selection methods and their application. Critical Reviews in Plant Sciences 3, 1–33.

Hayes, B.J., Visscher, P.M., Goddard, M.E., 2009. Increased accuracy of selection by using the realized relationship matrix. Genetics research 91, 47–60. URL: http://www.ncbi.nlm.nih.gov/pubmed/19220931, doi:10.1017/S0016672308009981.

Hazel, L.N., 1943. The Genetic basis for constructing selection indices. Genetics 28, 476–490. URL: http://www.genetics.org/content/28/6/476.abstract.

Hazel, L.N., Lush, J.L., 1942. The efficiency of three methods of selection. Journal of Heredity , 393–399.

Heffner, E.L., Lorenz, A.J., Jannink, J.L., Sorrells, M.E., . Plant breeding with genomic selection: Gain per unit time and cost. Crop science 50, 1681–1690. URL: http://cat.inist.fr/?aModele=afficheN&cpsidt=23164725.

Henderson, C., 1973. Sire evaluation and genetic trends, in: Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush, American Society of Animal Science and American Dairy Science Association, Champaign, IL. pp. 10–41.

Henderson, C.R., 1963. Selection index and expected genetic advance. Statistical genetics and plant breeding 982, 141–163.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423–447. URL: http://www.jstor.org/stable/2529430, doi:10.2307/2529430.

Hill, W.G., 1993. Variation in genetic identity within kinships. Heredity 71, 652. URL: http://dx.doi.org/10.1038/hdy.1993.190 http://10.0.4.14/hdy.1993.190.

Jacobsson, L., Park, H.B.B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, P.B., Andersson, L., 2005. Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. Genetical research 86, 115–125. URL: http://www.ncbi.nlm.nih.gov/pubmed/16356285%5Cnhttp://www.ncbi.nlm.nih.gov/ent doi:10.1017/S0016672305007767.

Laidig, F., Piepho, H.P., Drobek, T., Meyer, U., 2014. Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. Theoretical and Applied Genetics (TAG) 127, 2599–2617. doi:10.1007/s00122-014-2402-z.

Lerner, M., Cruden, D., 1947. The Heritability of Accumulative Monthly and Annual Egg Production. Poult Sci .

Longin, C.F.H., Mi, X., Würschum, T., 2015. Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theoretical and Applied Genetics 128, 1297–1306. URL: https://doi.org/10.1007/s00122-015-2505-1, doi:10.1007/s00122-015-2505-1.

Lorenz, A.J., 2013. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3 (Bethesda, Md.) 3, 481–91. URL: http://www.ncbi.nlm.nih.gov/pubmed/23450123, doi:10.1534/g3.112.004911.

Lush, J.L., 1937. Animal Breeding Plans. 3rd ed., Iowa State College Press, Ames, IA.

Lush, J.L., 1947. Family Merit and Individual Merit as Bases for Selection. Part I. The American Naturalist 81, 241–261. URL: http://www.jstor.org/stable/2457881.

Mackay, I., Horwell, A., Garner, J., White, J., McKee, J., Philpott, H., 2011. Reanalyses of the historical series of UK variety trials to quantify

the contributions of genetic and environmental factors to trends and variability in yield over time. Theoretical and Applied Genetics 122, 225–238. doi:10.1007/s00122-010-1438-y.

Maslow, A.H., 1943. A theory of human motivation. Psychological review 50, 370.

Massman, J.M., Jung, H.J.G., Bernardo, R., 2013. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Science 53, 58–66. doi:10.2135/cropsci2012.02.0112.

McGinnis, R.C., Shebeski, L.H., 1968. The reliability of single plant selection for yield in F2, in: Proceedings of the Third International Wheat Genetics Symposium, Canberra, Australia, pp. 109–114.

Mendel, G., 1866. Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereins Brünn 4, 3–47.

Meuwissen, T.H., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Meuwissen, T.H.E., 1997. Maximizing the response of selection with a predefined rate of inbreeding. Journal of Animal Science 75, 934–940. URL: http://dx.doi.org/10.2527/1997.754934x.

Peng, S., Laza, R.C., Visperas, R.M., Sanico, A.L., Cassman, K.G., Khush, G.S., 2000. Grain yield of rice cultivars and lines developed in the Philippines since 1966. Crop Science 40, 307–314. URL: http://dx.doi.org/10.2135/cropsci2000.402307x, doi:10.2135/cropsci2000.402307x.

Piepho, H.P., Laidig, F., Drobek, T., Meyer, U., 2014. Dissecting genetic and non genetic sources of long term yield trend in German official variety trials. Theoretical and Applied Genetics 127, 1009–1018. doi:10.1007/s00122-014-2275-1.

Piepho, H.P., Möhring, J., Melchinger, A.E., Büchse, A., 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161, 209–228. URL:

http://link.springer.com/10.1007/s10681-007-9449-8, doi:10.1007/s10681-007-9449-8.

Podlich, D., Cooper, M., 1998. QU-GENE: a simulation platform for quantitative analysis of genetic models. Bioinformatics (Oxford, England) 14, 632–653.

Poland, J.A., Bradbury, P.J., Buckler, E.S., Nelson, R.J., 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proceedings of the National Academy of Sciences URL: http://www.pnas.org/content/early/2011/04/07/1010894108.abstract.

R Development Core Team, 2016. R: A language and environment for statistical computing. URL: https://www.r-project.org/.

Rasmusson, D.C., Cannell, R.Q., 1970. Selection for grain yield and components of yield in barley. Crop Science 10, 51–54. URL: http://dx.doi.org/10.2135/cropsci1970.0011183X001000010020x, doi:10.2135/cropsci1970.0011183X001000010020x.

Rebetzke, G.J., Fischer, R.A., Van Herwaarden, A.F., Bonnett, D.G., Chenu, K., Rattey, A.R., Fettell, N.A., 2014. Plot size matters: Interference from intergenotypic competition in plant phenotyping studies. Functional Plant Biology 41, 107–118. doi:10.1071/FP13177.

Rutkoski, J., Singh, R., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J., Sorrells, M., 2015. Genetic Gain from Phenotypic and Genomic Selection for Quantitative Resistance to Stem Rust of Wheat. The Plant Genome 8. URL: https://www.crops.org/publications/tpg/abstracts/8/2/plantgenome2014.10.0074, doi:10.3835/plantgenome2014.10.0074.

Sheridan, A.K., 1988. Agreement between estimated and realized genetic parameters. Animal Breeding Abstracts 56, 877–889.

Smith, H., 1936. A discriminant function for plant selection. Annals of Human Genetics 7, 240–250.

University of Illinois Urbana-Champaign Department of Crop Science, 2007a. Data File : Means by year and generation of the oil and protein concentration measured each year (1896-2004). Technical Report. University of Illinois Urbana-Champaign. URL: http://hdl.handle.net/2142/3526.

University of Illinois Urbana-Champaign Department of Crop Science, 2007b. Data file : Raw data from each ear analyzed each year of the Illinois long-term selection experiment for oil and protein in corn (1896-2004). Technical Report. University of Illinois Urbana-Champaign. URL: http://hdl.handle.net/2142/3528.

Watson, A., Ghosh, S., Williams, M.J., Cuddy, W.S., Simmonds, J., Rey, M.D., Asyraf Md Hatta, M., Hinchliffe, A., Steed, A., Reynolds, D., Adamski, N.M., Breakspear, A., Korolev, A., Rayner, T., Dixon, L.E., Riaz, A., Martin, W., Ryan, M., Edwards, D., Batley, J., Raman, H., Carter, J., Rogers, C., Domoney, C., Moore, G., Harwood, W., Nicholson, P., Dieters, M.J., DeLacy, I.H., Zhou, J., Uauy, C., Boden, S.A., Park, R.F., Wulff, B.B.H., Hickey, L.T., 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. Nature Plants 4, 23–29. URL: https://doi.org/10.1038/s41477-017-0083-8, doi:10.1038/s41477-017-0083-8.

Weber, W., 1984. Selection in early generations, in: Lange, W., Zeven, A., Hogenboom, N. (Eds.), Selection in early generations. Proceedings of the 10th Congress of EUCARPIA, Pudoc, Wageningen. pp. 72–81.

Williams, E., Piepho, H.P., Whitaker, D., 2010. Augmented p-rep designs. Biometrical Journal 53, 19–27. URL: https://doi.org/10.1002/bimj.201000102, doi:10.1002/bimj.201000102.

Wright, S., 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. Proceedings of the National Academy of Sciences 6, 320–332.

Wright, S., 1921. Systems of mating. Genetics 6, 111–178.

Wright, S., 1930. Evolution in Mendelian populations. Genetics 16, 97–159. doi:10.1007/BF02459575.

Yabe, S., Iwata, H., Jannink, J.L., 2017. A Simple package to script and simulate breeding schemes: The Breeding Scheme Language. Crop Science 57, 1347–1354. URL: https://dl.sciencesocieties.org/publications/cs/articles/57/3/1347, doi:doi:10.2135/cropsci2016.06.0538.

9. Appendix: R-function used for simulation

The argument *nsim* indicates the number of simulations, *ngen* is the number of times identification of superior varieties is attempted, n is the number of selection units, *h2* is the heritability, *Vg* is the additive genetic variance, and *nsel* is the number of parents selected breeding cycle.

```
SimuSel<- function(nsim=10, ngen=3, n=100, h2=0.5, Vg=1, nsel=30){
  r<- sqrt(h2)
  p<- nsel/n
  i<- dnorm(qnorm(p))/p
  Ve<- Vg/h2- Vg
  stdG<- sqrt(Vg)
  stdE<- sqrt(Ve)
  succall<-c()
  for(j in 1:nsim){
    mn<- 0
    g<- rnorm(n,sd=stdG)
    e<- rnorm(n,sd=stdE)
    p<- g+e
    val0<- g[which.max(p)]
    succ<-c()
    for(i in 1:c(ngen-1)){
      mn<- mn+c(stdG*i*r)
      g<- rnorm(n,mean=mn, sd=stdG)
      e<- rnorm(n,sd=stdE)
      p<- g+e
      val<- g[which.max(p)]
      suc<- val>val0
      if(suc){
        val0<- val
      }
      succ<- append(succ, suc)
    }
    succall<- append(succall, !FALSE %in% succ)
  }
  probsuccess<- sum(succall)/nsim
  return(probsuccess)
}
```