Optimizing breeding schemes

Manual / Crossing strategies

Germplasm and trait introgression

- This manual provides guidelines for the introgression of traits and germplasm while maintaining closed recurrent selection programs.
- Published on 20/Jan/2020

Guidelines for germplasm and trait introgression

Author.

Giovanny E. Covarrubias-Pazaran // g.covarrubias@cgiar.org

Breeding Optimization Lead, CGIAR Excellence in Breeding Platform (EiB)

Editor(s).

Valentin Wimmer // KWS

Emily Ziemke // Corteva

J. Damien Platten // International Rice Research Institute (IRRI)

Sam Storr // EiB

BILL & MELINDA GATES foundation

CIMMYT MR
International Maize and Wheat Improvement Center

CGIAR
Excellence in Breeding Platform

Contents

Introduction

In plant breeding programs, the introduction of external germplasm is an important activity that helps to introduce new alleles for important or completely new traits. However, unless undertaken as a surgical activity with the proper considerations made, it is possible to have a negative impact on the ultimate goal of increasing the frequency of positive alleles. This is particularly the case in the common example of germplasm introduction taking place in the context of a forward breeding activity.

This manual is intended to outline general considerations to increase the frequency of positive alleles for important or new traits through the strategic introduction of external germplasm. Aimed at plant and animal breeders, it covers the following areas:

1. A clear definition of the different phases and stages involved in a breeding program, required to effectively identify areas where breeding strategies can be optimized.
2. The role of trait/diversity introgression in breeding programs and the considerations that have to be taken prior to considering trait introgression.
3. Recommendations for the design of trait/diversity introgression pipelines.

1. Breeding program phases

In order to critically assess the optimum application of a breeding strategy such as external germplasm introduction, it is important to clearly define the different phases and activities within a breeding program. While many competing definitions for breeding activities exist, CGIAR Excellence in Breeding Platform (EiB) methodologies rely on the understanding of a common terminology for effective knowledge exchange among breeders.

For the purpose of this manual and other EiB materials and activities, a breeding program can be divided into three distinct phases, according to the following definitions:

1. **Design phase.** Breeding program targets are defined. Target traits are identified and prioritized, and target genetic gains decided on. Ideally, this is documented in a product profile that takes into account consumer/market needs and breeding program capacity.
2. **Engineering phase.** The breeding program strategy is developed and put into place. This covers all breeding activities, such as trait/diversity introgression (also known as parent development), population improvement (also known as forward breeding), and product development (advanced testing of elite materials with commercial potential in farmer's conditions).
3. **Operations phase.** Elite materials are increased by agronomic services or the seed system, commercialization happens, and monitoring variety turnover occurs.

Dividing the breeding program in phases, i.e.: trait/diversity introgression, population improvement and product development is critical to take the right decisions. The population improvement phase requires closing the recurrent selection system.

The interactions that take place between key activities within each phase are described in **Figure 1**. In an optimal breeding scheme, trait introgression takes place between the design phase (ensuring it is guided by a clear strategy) and the engineering stage. The goal of trait/diversity introgression is to produce elite parents containing positive alleles that are either absent or present only in low frequencies in the core germplasm in forward-breeding / population improvement activities in the engineering phase. This makes it possible to introduce alleles that are new or required at greater frequencies while maintaining a closed population development system.

Three separate stages can be identified within the **engineering phase:**

1. **Trait/diversity introgression** (part of the breeding aiming to bring new alleles to the population improvement program through QTL/allele/haplotype deployment or in some cases in the final product through line augmentation).
2. **Population improvement** (part of the program aiming to increase the frequency of favorable alleles).
3. **Product development** (part of the program aiming to extract products from the population improvement sometimes complemented with some introgression).

While we will focus in the **Trait/diversity introgression**, the reader needs to remember that the final goal of this stage is to feed the population improvement stage with new alleles while disrupting the least possible.

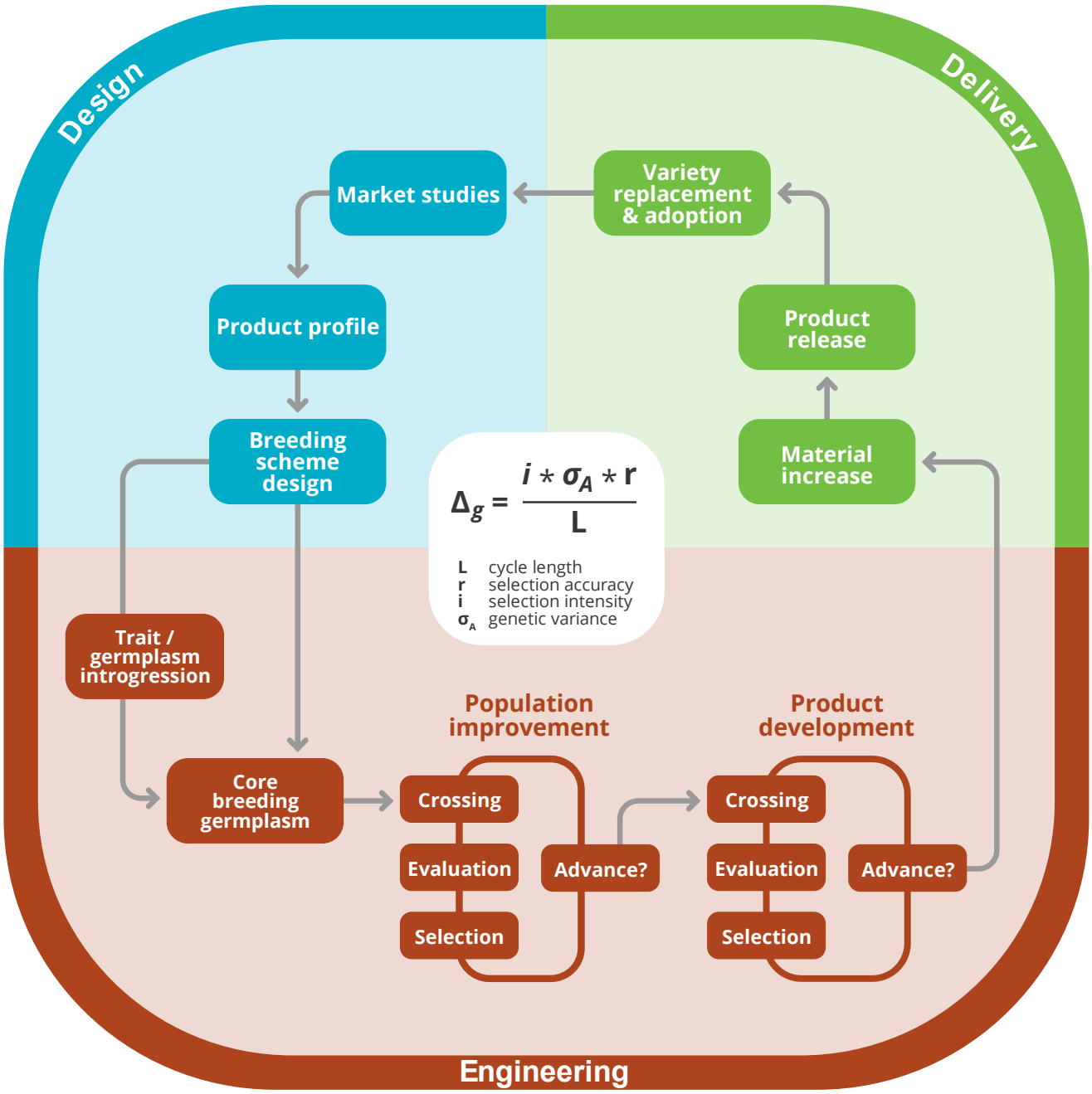

Figure 1. **Graphical representation of plant breeding as an industrial process.** In blue the design phase is displayed. The engineering phase comprises pre-breeding/trait-introgression, population improvement or forward breeding and initial product development (in orange). The operations phase (green) is focused on product delivery and its components are shown in blue.

> The best long-term strategy to deliver good products in the market is to increase the allele frequency of positive alleles for the traits of interest (through population-improvement/recurrent selection).

The goal of the population improvement stage is to increase the allele frequency of positive alleles for traits of interest in order to deliver good products to the market by using a recurrent selection methodology. The focus on positive allele frequency implies that once the best elite materials have been identified, it is crucial to close the system to non-elite germplasm and cross only elite by elite.

In a closed population development system, each cohort (cycle) of materials in evaluation and selection activities should originate from crosses between parents from previously evaluated and selected materials. In an optimized breeding program, ~90-100% of new parents in each generation of the population improvement stage should come from within the recurrent selection program, unless there is another strategy to keep the system closed. The small number of remaining parents can come from the trait/diversity introgression pipeline as long as certain considerations are taken, as outlined below.

2. Considerations for germplasm and trait introgression

The introduction of new material in the engineering phase of the breeding program, while an important source of alleles, can have an important impact on the overall strategy to increase the frequency of positive alleles for traits targeted by the breeding program. For this reason, special care should always be taken when introducing any new material in the population improvement component and managing impact on the general breeding strategy.

Figure 2 demonstrates improper introgression can have a negative impact. In this example, the breeding program started with an initial diploid forward-breeding population with 100 parents that segregate for a single disease resistant gene with an allele frequency of p=0.2, meaning that there are 40 alleles "+" (p=0.2) in the population and 160 "-" alleles (q=0.8). After four generations of recurrent selection, the allele frequency of the positive allele is increased to p=0.6 (120 "+" alleles in the population). At this point, if the breeding program were to introgress 50 donor parents obtained from other sources or programs where this disease resistance gene is not of interest or relevant, the allele frequency in these 50 new individuals may be as low as p=0.05 (10 "+" alleles). In such a case, the merging of populations will create 150 parents with p=0.41, considerably reducing the allele frequency of the gene of interest potentially undoing generations of recurrent selection (and consequently, time and resources).

The same consideration applies to quantitative traits. Introgressing new germplasm into parents for a new cycle without first assessing their breeding value (i.e. for yield) and, more importantly, without knowing the total genetic merit of those donors for all the traits (i.e. based on a robust selection index) has negative repercussions in the allele frequencies of the traits of interest just as demonstrated for a simple trait in **Figure 2**.

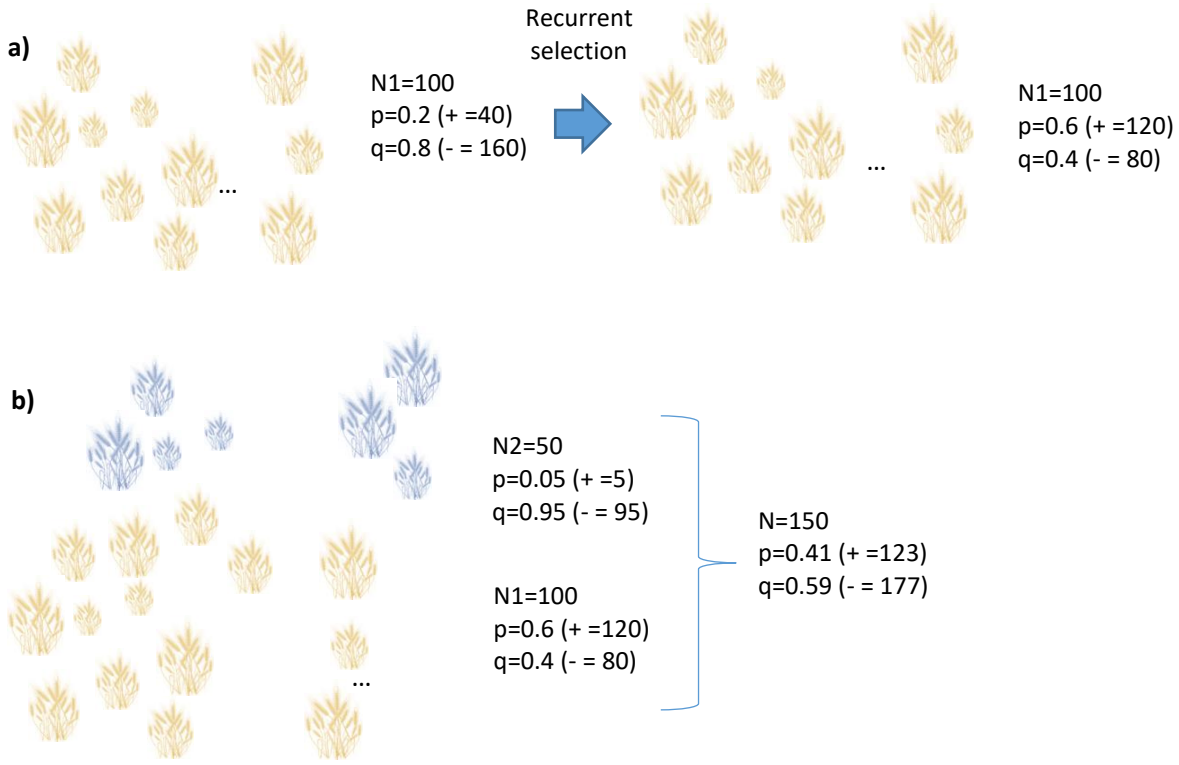

Figure 2. Example of the effect of foreign germplasm introduction in forward breeding. In a) the effect of recurrent selection in the increase of allele frequency for a trait governed by a single gene is displayed. In b) the decrease of allele frequencies for a positive allele as a consequence of foreign germplasm introduction is shown.

The introgression of germplasm to be used as a parent without knowing its breeding value (i.e. for yield) and the total genetic merit of those donors for all the traits (i.e. based on a robust selection index) has negative repercussions for the allele frequencies of the traits of interest.

2.1 Tools and recommendations for effective trait/diversity introgression

Effective trait introgression is a prerequisite to successful population improvement, and therefore the ability to develop and deliver the desired product. In order to ensure the optimum and congruent execution of these three stages, it is best practice to clearly delineate and document them in the breeding program, with the option of using visual tools to generate transparency in the breeding team.

Flowcharts are a common and effective way to document processes, in which tasks and decisions are mapped as defined steps and decision points. In the case of trait introgression, there are several possible tasks and decisions that are part of this process that can be defined in a logical order, as shown in **Figure 3.** This generic diagram can be used as a starting point for breeding programs to adapt to their own task structure, and can be developed using commonly available commercial or freeware applications.

In an optimized breeding program, ~90-100% of the new parents of the next generation should come from the recurrent selection program, unless there is another strategy to keep the system closed. The rest can come from other sources as long as there is a clear understanding of their genetic merit.

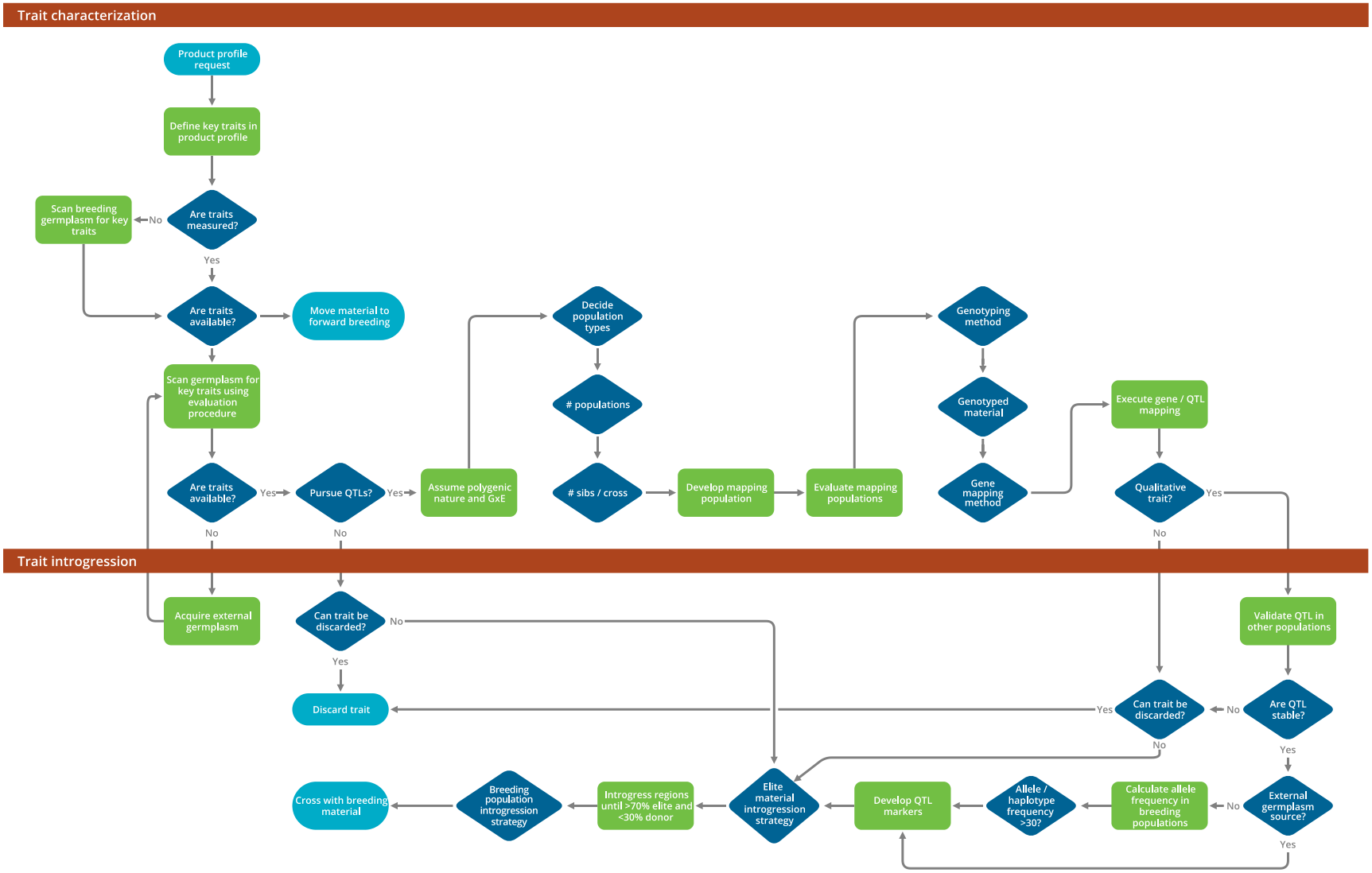

Figure 3. Example of the use of flowcharts to document tasks and decisions in a breeding program. An example of a trait-introgression pipeline is depicted dividing the introgression in 1) trait characterization followed by 2) trait introgression.

The introduction of new traits and diversity using external germplasm can be a valuable activity, but must be undertaken considering potential repercussions for the allele frequencies of the traits of interest. In particular, the introduction of material with unknown breeding value or genetic merit can negatively impact allele frequencies and the population means of the traits of interest, and consequently waste time and resources.

There are two main options available to guarantee the genetic merit of foreign parents:

1. Extensive testing of foreign parents alongside the breeding population.
2. Introgression of desired traits in the most elite material available (i.e. the latest selected material in the recurrent selection program), in order to exclusively use elite material in population improvement.

2.2 The germplasm and trait/diversity introgression process

The trait/diversity introgression is usually tackled either 1) with the purpose of increasing the frequency of the new alleles in the breeding population, or 2) with the purpose to introduce a trait into an elite material or product (line augmentation).

The first type of introgression, the one that has the purpose of increasing the frequency of the new alleles in the breeding population, usually requires the following activities to take place:

1. Development or availability of a product profile and key traits to breed for.
2. Assessment of the availability of traits in the population improvement germplasm.
3. In case the trait(s) are not available in the population, identification of additional sources.
4. In the case of simple traits, one option is gene-mapping to extract quantitative trait loci (QTL) markers and move to production to make backcrossing more efficient. Such QTLs (few so they are manageable) should explain more than 30% of the genotypic variation and must be stable in multiple backgrounds.
5. Elite germplasm (upper tails) coming from the latest recurrent selection cycle should be used as recurrent parents to introgress the trait(s) of interest from the donor parent(s), using a backcrossing approach (**Figure 4**), until attainment of over 87.5% (BC3) of the recurrent parent or until it is ensured that the genetic merit of the new material is above the mean genetic merit of our population.

In the case that the material containing the trait of interest is to be used directly in population improvement, it must be evaluated alongside the breeding population to ensure that any material used has a total genetic merit higher than the average total genetic merit of the breeding population.

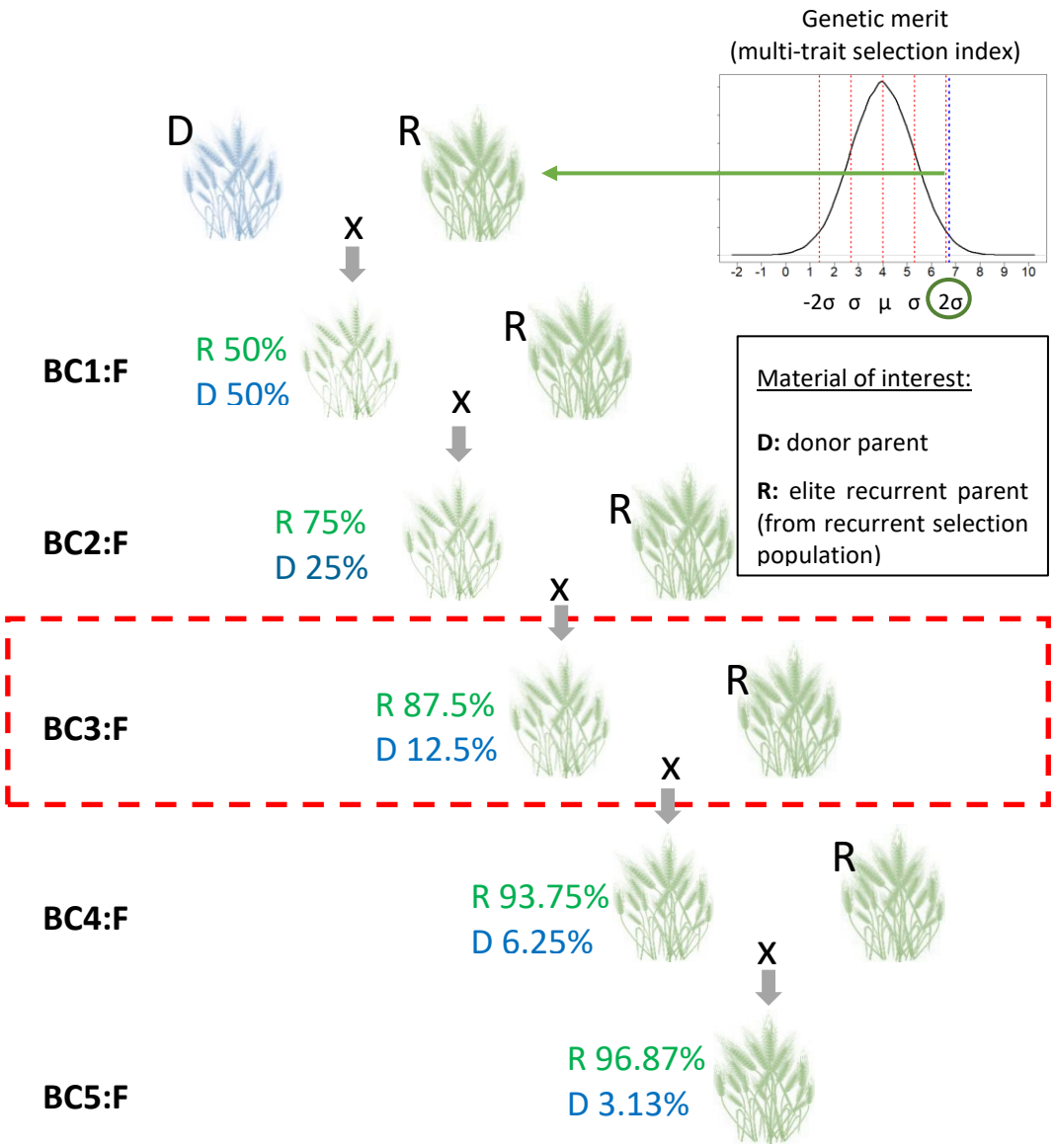

Figure 4. **Example of classical backcrossing to introduce a trait of interest in a breeding population without affecting the recurrent selection population with the purpose of increasing the allele frequency of positive alleles.** It is recommended to use elite parents from the upper tail of the breeding population based on a selection index for total genetic merit (i.e. two standard deviations) from the population improvement population to introduce a new trait through backcrossing (until BC3 if the parents falls 2σ, and BC4 if the parent falls 1σ from the mean total genetic merit; if MABC is pursued this could be earlier). Once the trait has been introduced in an elite background this material can be used as parent in the new generation to increase the allele frequency of the new trait in the breeding population.

On the other hand, the introgression with the purpose to introduce a trait into an elite material or final product, also called line augmentation, is faster and far cheaper (per background), but relies on high-quality donor material being available (**Figure 5**). A good review of the requirements and process of line augmentation can be found at Cobb et al. (2019).

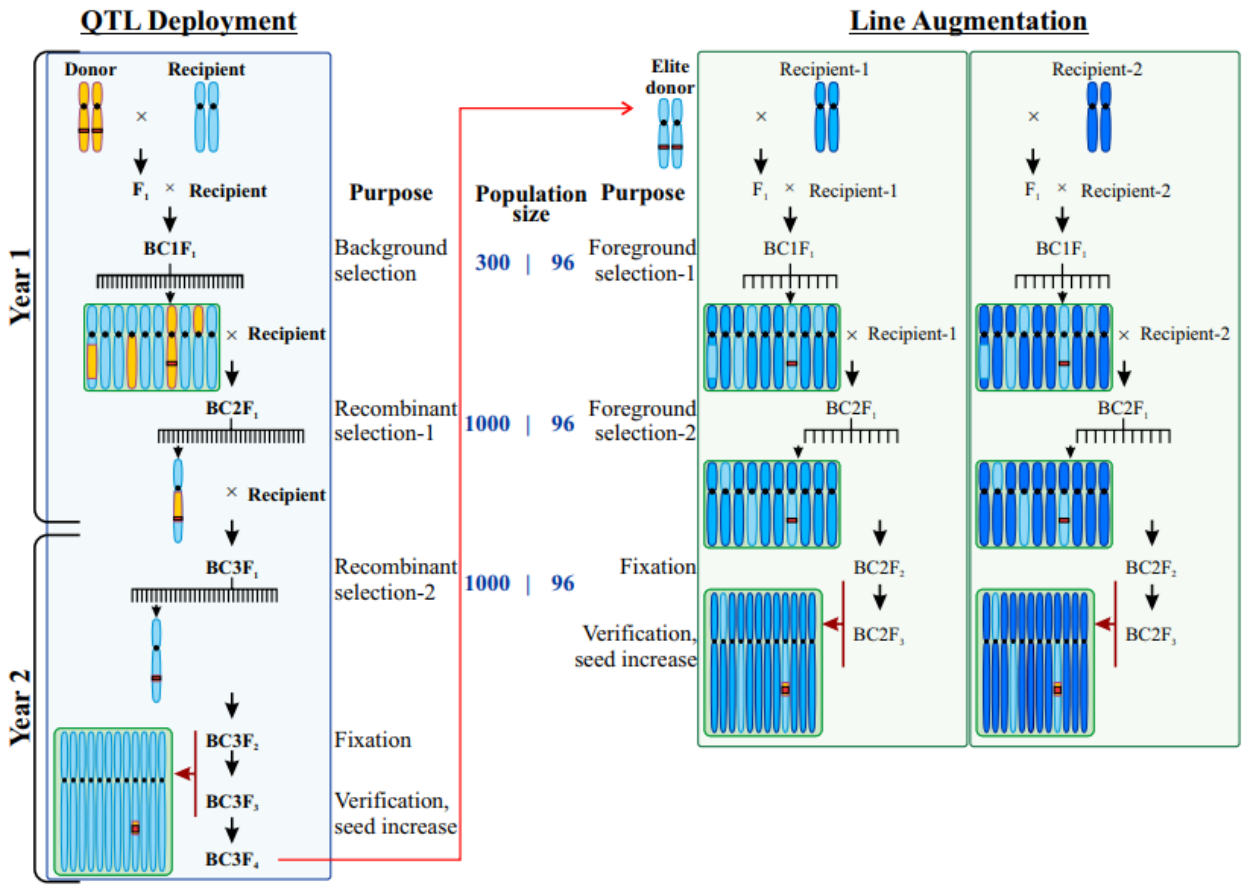

Figure 5 **Example of the contrast of workflows between regular allele/QTL/haplotype deployment and line augmentation**. The focus of QTL deployment on producing quality introgressions necessitates large populations for recombinant selection and advanced backcross generations to clean up the genomic background of poor-quality donor landraces. Line augmentation starts with a high-quality donor from QTL deployment and introduces the new locus into additional elite backgrounds, with only a few backcrosses and no recombinant selection. This means augmentation is faster and far cheaper (per background), but relies on high-quality donor material being available. Figure and text taken from Cobb et al. (2019)

2.3 Maintaining diversity in breeding programs

It is a common practice for breeding programs to make wide crosses, based on the belief that genetic diversity needs to be expanded rather than an understanding of quantitative genetics principles or knowledge of how diversity changes in a breeding program. As a result, the allele frequency of important traits can be negatively impacted.

Empirical evidence and simulations have shown that exhausting diversity for a trait is very difficult when selection intensities are low to moderate, even for programs based on very small populations [i.e. formed of 12-20 parental units (i.e. lines or families) breeding a couple of traits]. For example, the Illinois long-term selection experiment undertaken more than 100 cycles of selection (selection intensities of 0.64 to 1.05, equivalent to selecting 60% to 35%, and up to 20% of elite materials on population sizes of 60-120 materials generated initially when the program started out of 12-30 parents), maintaining genetic gains to date accounting for more than 20 standard deviations from the original population mean (Moose et al. 2004, Dudley et al. 1974). A proposed explanation is an intermediate-high number of gene (~100-200), mutations driven by selection, and change of epistatic landscape as genes become fixed (Dudley 2007).

Although, simulations cannot encapsulate the full complexity of diversity maintenance in the long-term, this technique can validate the theory of small increases in frequency of favorable alleles at hundreds of loci. For example, using the initial mean, variance, heritability and number of genes behind the *oil% content* trait in the Illinois long-term selection experiment, it is possible to simulate the expected genetic gain and change in variance after 100 cycles of selection (**Figure 6**). The results show clearly that realized and expected genetic gain can be simulated to reflect the observed reality, although the epistatic and mutational effects are more difficult to include. Exhausting genetic diversity is possible in simulation while in reality variation doesn't disappear completely given other genetic mechanisms. As it is difficult to exhaust diversity for a trait when selection intensities are low to moderate, even for very small populations, the practice of uncontrolled diversity introduction should therefore be discouraged.

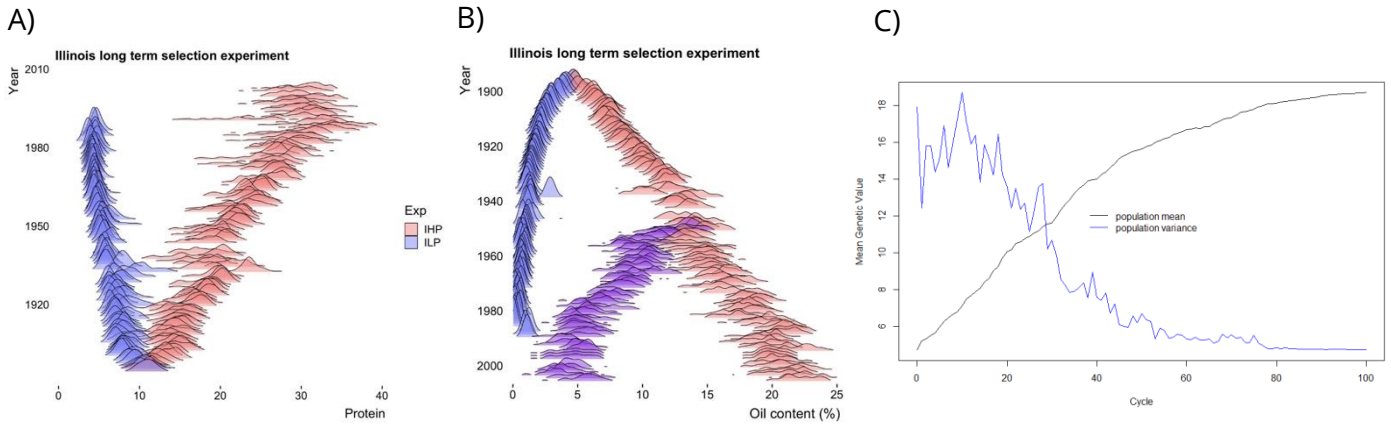

Figure 6. Results of the Illinois long-term selection experiment (depicted by Graham Coop. GitHub link in Literature section) and validation of observations based on simulations. In A) the change in population mean and variance for protein (IHP: Illinois high-protein; ILP: Illinois low-protein) along more than 100 cycles of RS are displayed. In B) the same process for %oil content is displayed. In C) the expected genetic gain (black line) and variance (blue line) exhausting trends for oil % content based on an initial mean, variance and number of genes (200 assumed according to Dudley 2007).

Practical examples and simulations have shown that exhausting diversity for a trait is very difficult when selection intensities are low to moderate even in programs based on a reduced number of parental units (i.e. populations formed by 12-20 parents for few traits of interest).

Recommendations for maintaining diversity without breaking the allele frequencies in a recurrent selection program can be summarized as follows. These are only general guidelines that should be optimized for your specific situation:

1. Make sure that initial/core breeding population contains all the traits of interest (defined in a product profile) and that appropriate variation exists [i.e. at least 12-20 selection units (i.e. lines, S_0 families, etc.) with high breeding value for a couple of traits defined in the product profiles, and no more than 100 total for complex product profiles].
2. Initiate the recurrent selection process by closing the system to non-elite germplasm, avoiding more than 5-10% of parental units coming from outside the program (unless another strategy to keep the system closed is in place). If introduction is necessary, use the backcrossing scheme proposed in **Figure 4**. Use a selection index for total genetic merit to select the new parents for all traits.
3. Keep a healthy number of parental units each generation [i.e. > 10% for few traits. If product profile prioritize > 5 traits select no more than 75-100 parental units (i.e. lines, S_0 families, etc.) every cycle representing 5 to 50% of the top selection units of the distribution]. Maintain a selection intensity between 0.8 and 2.06, corresponding to selecting from 50% to 5% of the top individuals depending how aggressive a scheme is adopted to accelerate the breeding process (more aggressive schemes will exhaust diversity quicker).

One final consideration regarding the number of parental units used for crossing to form the new generation is that, for each trait of interest, the number of parental units does not need to be extremely high. Simulations have shown that, given a fixed number of plots and crosses (i.e. 100 crosses with 20 progeny to evaluate 2000 individuals for a single trait), the number of parental units behind those crosses need not be greater than 30 for a few quantitative traits defined in the product profile if the window of the breeding program aims to be less than 25 cycles of recurrent selection, as they will produce almost identical genetic gains (**Figure 7**). A rule of thumb of using as many parents as targeted breeding cycles has been proposed (Bernardo, 2010).

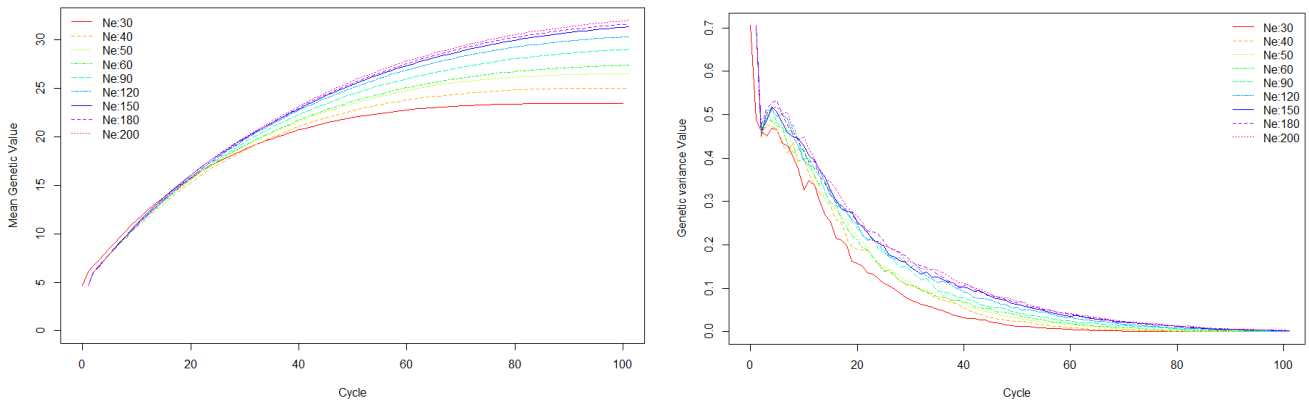

Figure 7. Effect of the different numbers of parents in the genetic gain and variance when the number of crosses (100), progeny (20) and plots is fixed (2000). The different lines represent a different number of parents used to form the new generation. In a selection window of 25 selection cycles, having more than 30 parents has little effect on the genetic gain and variance for a single trait, but following that window the number of parents has an effect on the long-term genetic gain. Ne refers to effective population.

In a selection window of 25 cycles having more than 30 selection units (i.e. lines or families) has little effect in the genetic gain and variance for a single trait but going after that window the number of parents has an effect on the long-term genetic gain.

2.4 Special considerations for small programs

When a program uses a small number of parents (i.e. ~20-30) each selection cycle of new parents, the question of how many donor parents can be introduced may arise. As mentioned, a breeding program should take care not to include more than 5-10% of parents from trait introgression pipelines. The actual number of donors is adjusted for program size, but the 5-10% rule should be respected (unless another strategy to increase the frequency of favorable alleles is in place). A robust breeding program will conduct both activities, introgressing traits and new allelic diversity in elite backgrounds in a controlled manner to later increase the frequency of favorable alleles in the population improvement program.

2.5 Special considerations for hybrid crops

In the case of hybrid crops, where the main focus is on selling/offering heterosis, the population improvement strategy is focused on increasing not only the frequency of favorable alleles with an additive effect, but also the heterosis (non-additive interactions) by splitting the genetic material in genetic groups called heterotic groups. Special considerations should be made for trait introgression in this context, beyond those outlined above.

The most important consideration is that any material introduced with the goal of introgressing a trait or new allelic diversity *should not break the heterotic patterns*. Heterotic groups are usually the result of natural or provoked divergent evolution into different directions in which some groups are *complementary* to each other for traits of interest. For example, in **Figure 8** an initial population is arbitrarily divided in 2 pools followed by 9 generations of reciprocal recurrent selection (RRS)[1] which results in a clear generation of pools that will be complementary to each other as long as each pool uses material from the opposite pool as testers to ensure *complementarity*.

[1] For more information on RSS, see the EiB manual on Reciprocal Recurrent Selection.

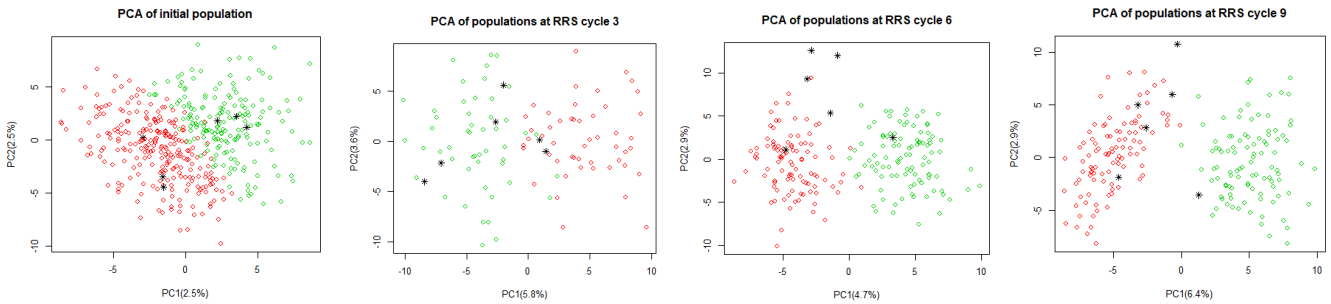

Figure 8. Evolution of heterotic pools as a consequence of reciprocal recurrent selection (RRS). From left to right, an initial population is arbitrarily divided in two groups and testers are selected, then nine generations of RRS are applied resulting in the creation of heterotic groups. Divergence will happen as long as pools are kept separated, the proper increase of heterosis among the pools will depend from the selection based on general combining ability (GCA) among pools.

The most important (additional) consideration for hybrid crops is that any trait or variation introduced should not break the heterotic patterns.

In order to avoid breaking the heterotic patterns, each time that new allelic diversity is planned to be introduced, it is necessary to ensure the right classification of the new material through either of the following three approaches (**Figure 9**).

1. The new material should be crossed against a couple of representative materials from each heterotic group (elite material of preference to initiate the backcrossing approach as quick as possible after the classification) and the progeny evaluated. Based on the performance of the progeny [specific combining ability (SCA) displayed] we can define to which heterotic group the new material belongs to. For example, if the cross of the new material yields better SCA progeny when crossed against pool 1, it means that the new material belongs to pool 2. The same logic applies if the result is opposite.
2. Both pools can be genotyped with enough genetic markers to create a principal component analysis (PCA) plot or other multivariate discrimination/classification methodology to understand the genetic constellations. The new material should be genotyped and plotted together with the heterotic pools. Based on its location, the new material should be classified into the pool that is closer to it.
3. If both pools have a detailed pedigree, this may indicate the separation among the pools and allow new materials to be traced to each pool. The new material introduced or acquired should have a detailed pedigree record that allows the classification of the material into one of the existing pools.

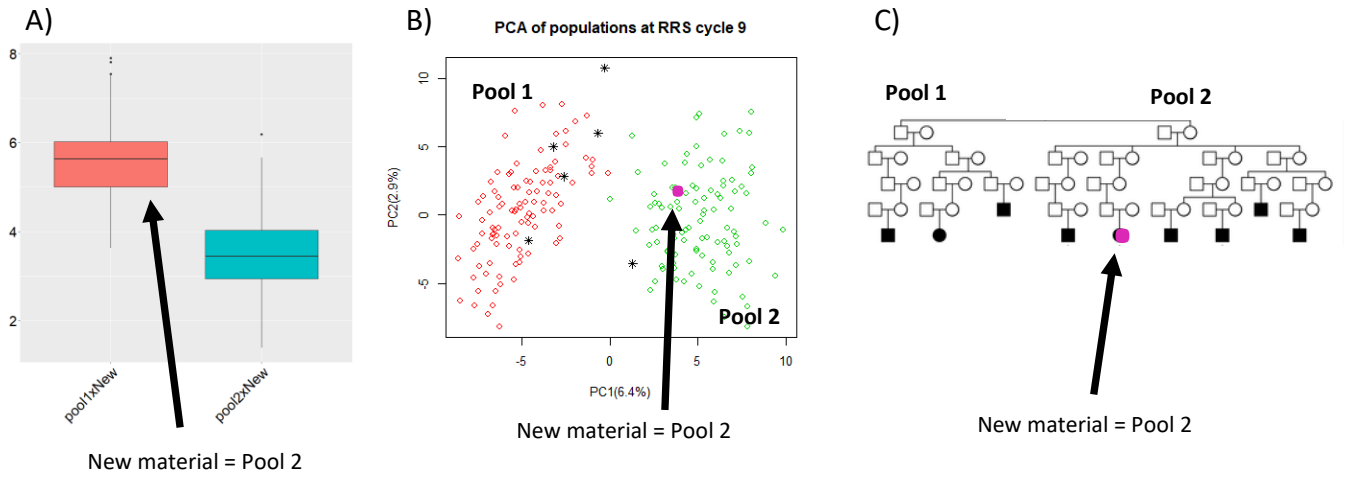

Figure 9. Three approaches for introducing new material in a hybrid population. In A) the new material is crossed by a representative sample of each pool and progeny are evaluated providing the classification depending on which cross provide more heterosis. In B) the new material together with the parents of the current pools are genotyped and a PCA plot is run to see where the material falls in the genotype constellations, the classification is based to proximity. In C) the breeding material and the new germplasm possess a detailed pedigree that allows the classification of the new material.

Conclusion and recommendations

This manual outlined how to clearly distinguish population improvement and trait/diversity introgression in a breeding program in order to more clearly understand the role of trait/diversity introgression and the guidelines that must be taken to successfully perform this activity. Additional considerations for trait and germplasm introgression in hybrid programs was also presented. Finally, it was shown how to use flowcharts as a visual tool to document breeding processes and assist in the design of an optimal trait introgression pipeline.

References

1. Bernardo, R., 2002. Breeding for quantitative traits in plants (Vol. 1, p. 369). Woodbury: Stemma Press.
2. Cobb, Joshua N., Partha S. Biswas, and J. Damien Platten. 2019. Back to the future: revisiting MAS as a tool for modern plant breeding. Theoretical and Applied Genetics 132.3: 647-667.
3. Dudley, J. W., and R. J. Lambert. 2004. 100 generations of selection for oil and protein in corn. Plant breeding reviews 24.1, 79-110.
4. Falconer, D.S., and T.F.C. Mackay. 2005. Introduction to quantitative genetics. 4th ed. Pearson Prentice Hall, Upper Saddle River, NJ.
5. Coop, G. https://github.com/cooplab/popgen-notes/tree/master/Journal_figs/Quant_gen/Illinois_long_term_selection_corn
6. Moose, S., Dudley, J., and Rocheford, T. 2004. Maize selection passes the century mark: a unique resource for 21st century genomics. Trends in Plant Science 9:7, 358-364.
7. Mrode, R., 2014 Linear models for the prediction of animal breeding values. 3rd ed. CABI Press, Oxfordshire OX10 8DE UK.
8. Smith, A., Cullis, B., and Gilmour, A. 2001. The analysis of crop variety evaluation data in Australia. Aust. N. Z. J. Stat. 43:129–145. doi:10.1111/1467-842X.00163
9. Walsh, B., and Lynch, M. 2018. Evolution and selection of quantitative traits. 1st ed. Oxford Univ. Press, Oxford, UK. doi:10.1093/ oso/9780198830870.001.0001