# Breeding: Plants, Modern

**JB Holland,** North Carolina State University, Raleigh, NC, USA

Published by Elsevier Inc.

## Glossary

**Epistasis** The dependence of one gene's effects on the state of a different gene in the same organism.
**Genomic selection** The use of a large number of genetic markers across the genome to predict breeding values of plants that have not been grown and tested yet.
**Germplasm** The material representing the genetic diversity available to a breeding program, often stored as seeds or vegetative plant cuttings or maintained in living nurseries.
**Linkage disequilibrium** Alleles at different genes that cooccur in individuals more often or less often than expected by chance are in linkage disequilibrium; more precisely this is termed 'gametic phase disequilibrium' because it does not require physical linkage on a common chromosome.
**Marker-assisted selection** Selection for deoxyribonucleic acid markers previously identified as linked to or inside of desired alleles at genes controlling important traits.
**Quantitative trait loci (QTL)** A chromosomal region associated with variation for a quantitative trait.

## Historical Development of Plant Breeding Methods

### The Diversity of Mating Systems in Plants

Mating systems in plants are highly diverse, much more-so than in vertebrate animals. Some plant species have exclusively or primarily outcrossing ('allogamous') mating behavior enforced or promoted by sex chromosomes (like domesticated animals), self-incompatibility mechanisms, or spatial/temporal separation of male and female flower development. Typically, these species have morphological and developmental features that represent adaptations to insect and wind-borne pollen dispersal. Primarily outcrossing species tend to have population structures similar to many animal species, characterized by higher levels of genetic diversity within populations, higher levels of heterozygosity within individuals, and more genetic load (recessive alleles that reduce fitness when homozygous). Many perennial tree and forage species have outcrossing mating systems and population structures (Richards, 1997), as do most fruit crops (Janik, 2006). Many annual vegetable crops and some cereals such as maize are also predominantly outcrossing (Allard, 1960).

In contrast, many other important crop species are primarily self-fertilized ('autogamous') annual species. Many self-fertilized species have morphological/developmental features that promote self-pollination (such as maintaining closed flowers until after pollen shed) and often lack the 'showy' floral features of insect/animal pollinated species. Self-fertilized species tend to have significantly lower levels of genetic variation within populations, lower heterozygosity within individuals, and generally suffer little inbreeding depression, presumably because much of their genetic load was purged during the evolution of the self-fertilization mating system. Most of the cereal grass crops (such as wheat and rice) and annual seed legume crops (such as soybean, common bean, pea, and peanut), and a number of other important crops (including tomato, tobacco, cotton, lettuce, and pepper) are predominantly self-pollinated (Allard, 1960).

Still another group of plants reproduces using both asexual and sexual means (to varying proportions). Important crops in this category include potato, sweet potato, sugarcane, and many perennial grasses (Ortiz *et al.*, 2006; Wu *et al.*, 2006). Many ornamental cultivars are sterile interspecific hybrids and are necessarily propagated asexually. Asexual propagation can occur naturally by spreading stem structures, such as rhizomes, or by asexually produced seed ('agamospermy' or 'apomixis'). Apomixis occurs when viable seeds are formed in the absence of sexual reproduction (Richards, 1997). Beyond naturally occurring mechanisms for asexual reproduction, humans have devised methods to propagate vegetative plant parts (by stem or tuber cuttings, grafting, or even tissue culture). Grafting has deep roots (so to speak) in agriculture and was known to the ancients (Janik, 2006). Citrus is an interesting example that demonstrates the reproductive flexibility that even a single plant group can possess. Citrus trees are naturally self-pollinated but also can be outcrossed easily, usually forming seeds from sexual reproduction. Some types of Citrus, however, will also produce seeds asexually, which represent clonal copies of the maternal plant. Some forms of Citrus will produce both sexual and asexual embryos within the same seed (Richards, 1997). Beyond this natural variability in the reproductive mode of citrus, plant breeders have devised methods of asexual propagation from clonal propagation of a genotype by grafting vegetative cuttings to artificially producing novel hybrids by cell fusions (Gmitter *et al.*, 2009).

These mating system and cultivar type divisions do not follow phylogenetic relationships; all types may (and often do) occur within a taxonomic family (Richards, 1997). For example, the grass family, *Poacea*, contains cross-fertilized, self-fertilized, and asexually reproducing groups (Richards, 1997). Furthermore, most perennial grass species can be propagated vegetatively, and some grasses also reproduce apomictically. For plant breeders, the reproductive behavior of a species is almost always the critical determinant of the breeding method that will be most commonly employed for crop improvement (Allard, 1960). Thus, breeders of soybean (a member of the

eudicot family) and spring-sown wheat (monocot family) will use much more similar breeding methods than will wheat and maize (both monocots) breeders. Perhaps more dramatically, breeding methods used for some forest tree species (mostly outcrossed, and with substantial attention paid to measurements of individual plants and their known pedigree relationships with other trees) are fundamentally more like animal breeding methods than wheat breeding methods.

Furthermore, in the same way that plant reproductive systems largely determine the genetic structure of natural populations, the reproductive system also is the most important factor in determining the predominant genetic structure of cultivars. The mating system has two major impacts on cultivar type. First, the mating system has a critical effect on the relationship between heterozygosity and vigor in the species (Allard, 1960; Richards, 1997). Primarily outcrossing species rely on heterozygosity (having two distinct copies of a gene) to shield the plant from the effects of deleterious recessive mutations. Thus, most outcrossing species exhibit strikingly reduced vigor and fitness when inbred (inbreeding depression) and greater vigor when crossed to unrelated plants of the same species (hybrid vigor). In contrast, self-fertilizing species have adapted to a genetic constitution characterized by high homozygosity, with unfavorable and lethal mutations consistently exposed phenotypically. Self-fertilization appears to have evolved from cross-fertilization in many plant lineages when the detrimental effects of inbreeding were outweighed by the advantages of not having to rely on a separate plant for pollen, as could be important in colonization of geographically isolated habitats, for example (Stebbins, 1974). The second major impact of mating system on cultivar types is its effect on the economic efficiency of cultivar propagation and dissemination. For example, even predominantly self-fertilized species may exhibit some hybrid vigor, so F1 cultivars could be an optimal genetic structure in terms of vigor. Yet, these species also tend to have floral structures and pollination behavior that maximizes self-fertilization and makes cross-fertilization difficult. For example, whereas naturally outcrossing maize has separate male and female flowers facilitating easy emasculation for hybrid seed production, manual emasculation of self-fertilizing species is very time-consuming and requires significant experience to perform well without damaging the female organs. In some species, this has been overcome to some extent by genetic or chemical male-sterility systems, but even then, these species tend to shed relatively little pollen, so that manual pollination and a larger proportion of male to female parent plants may be required to produce seed, increasing seed production costs (Li and Yuan, 2010; Wilson, 1984).

In summary, plant breeding methods and cultivar types can be roughly characterized according to the predominant reproductive system of the species. The result of this is that the genetic structure of cultivars can be categorized according to the level of heterozygosity (within-plant genetic diversity) and heterogeneity (among-plant genetic diversity; **Figure 1**). **Figure 1** is an oversimplification, but captures the large pattern of genetic population structure in different cultivar types. Many older landrace varieties (maintained on farm by farmers over long periods of time) contain substantial genetic diversity. In the case of outcrossing species, this diversity is a natural result of the genetic recombination that occurs with each generation of seed production. In the case of self-fertilizing species, diversity is a result of a mixture of seeds from distinct pure lines resulting from recurrent inbreeding spiked by occasional outcrossing to generate new variability (Allard, 1999). Much of the drive of modern plant breeding has been toward uniformity within cultivars, achieved by isolating pure-breeding lines in the self-fertilized species, or by creating uniform but highly heterozygous F1 cultivars in some cross-pollinated species. As mentioned already, some predominantly inbreeding species are

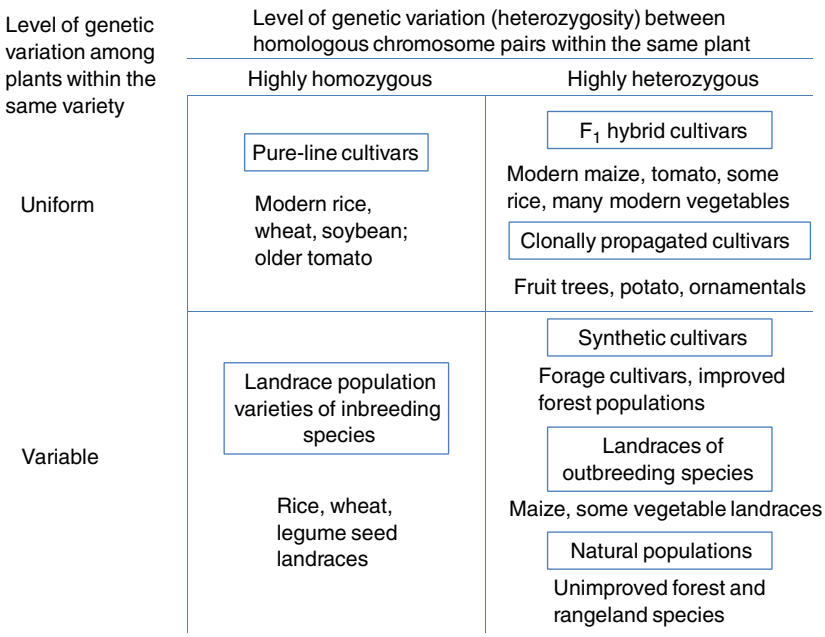


**Figure 1** Classification of cultivar types according to genetic diversity among plants within a cultivar and heterozygosity within a plant.

represented by modern F1 cultivars despite the difficulties in seed production, as a way to capture hybrid vigor (which may be limited) and, perhaps more importantly, to more easily combine favorable attributes of different parental lines (e.g., by combining disease resistance with specific fruit quality characteristics).

As with breeding methods, the cultivar types are largely determined by the mating system not the taxonomic grouping of the species, which is why **Figure 1** does not follow a phylogenetic pattern. The predominant genetic structure in most cereal grass cultivars like wheat is a highly homozygous and highly homogenous pure-line variety, whereas maize (an outcrossing cereal grass) cultivars are most often highly heterozygous F1 hybrids.

## The Concept of Heritability

Ancient plant domesticators and breeders achieved tremendous gains from selection with no understanding of the biological mechanisms underlying inheritance. Key to this success, however, was the understanding that offspring tend to look more like their parents than unrelated individuals. Thus, it was surely intuitive for ancient plant breeders to use seeds from more desirable plants to sow the next generation. Much less intuitive was Darwin's insight that natural selection, if repeated over many generations, could result in the appearance of new species (Darwin, 1859). Darwin developed his theory of the origin of species by natural selection with an entirely incorrect understanding of the mechanisms of inheritance (Darwin, 1859). Nevertheless, he understood correctly that most traits related to fitness are to some extent heritable from parent to offspring. Thus, absent an understanding of genetics, the concept of inheritance alone is sufficient for effective selection and for basic understanding of the evolution of species.

The synthesis of Darwinian concepts of evolution by natural selection with Mendel's genetics greatly enriched our understanding of both evolution and genetics. The recent developments in genome biology have similarly had profound impacts on these areas of science. Modern plant breeding is a naturally integrative applied field of science, whereby breeders leverage the understanding of inheritance, trait expression, and population dynamics to design more efficient and more effective means of crop improvement. Similarly, plant breeders apply tools and technologies (from genomics, computer science, statistics, agricultural engineering, and so forth) to the extent that their application will help make selection responses better for a fixed input of time and resources. Thus, whereas a simple understanding of inheritance enabled dramatic changes in crops over thousands of years, modern breeders hope that better understanding of the specific details of the inheritance of important traits and the application of tools that permit more accurate discrimination among the potential utility of different plants will greatly speed up the rate of gain from selection. Indeed, recent history suggests that plant breeders have achieved more dramatic gains in the last 100 years in some crops than occurred previously over much longer time scales. The changes in yield potential, stress tolerance, and population genetic structure of maize represent a good example of this phenomenon (Duvick *et al.*, 2004; van Heerwaarden *et al.*, 2012).

A major step toward modern plant breeding was achieved by Johannsen, who studied the inheritance of bean size and weight (Allard, 1960; Johannsen, 1903, 1911), demonstrating that bean populations were a mixture of distinct pure lines, and that selection among those lines would result in heritable changes. Importantly, he noted that while there was also variation among plants within a pure line, selection within lines had no effect on the progeny mean values. He realized this was because there are two components of observable 'phenotypic' variation: (1) 'genotypic' variation because of differences in genetic composition, and (2) variation because of environmental factors. Within a pure line, there is almost no genotypic variation; thus the observed variation is almost entirely because of different reactions of a common genotype to varying environmental conditions. Differences between pure lines are because of both genotypic differences and also environmental differences, so some proportion of those differences will contribute to a response to selection in the progeny. This distinction between genetic and nongenetic causes of variation was critical to providing a sound scientific basis for selection. It made clear that different population structures had different levels of genetic variation, and that selection would have more effect in population structures with more genetic variation. Additionally, by distinguishing the important contribution of environmental effects to observable phenotypes, Johannsen's insights led breeders to consider more carefully how to reduce the contribution of environmental differences among plants under selection, so that selections were determined more by genotype variation rather than nonheritable environmental variation.

Following critical early experiments demonstrating that continuously varying traits could arise as the result of the combined action of multiple genes and the environment (East, 1910, 1916), Fisher (1918) formulated a statistical genetic model that partitioned the overall effect of genotypes into the 'additive' (average) effects of their component genes, dominance interactions between the two copies (alleles) of the same gene in an individual, and epistatic interaction between alleles at different genes. This model led to a formulation of the genotypic variance in a population as the sum of the variation because of additive, dominance, and epistatic effects of component genes.

The importance of conceptually distinguishing the different components of heritable genotypic variation was emphasized by Lush (1945), who coined the term 'heritability' to refer to the proportion of total phenotypic variation observed for some trait in a population that is heritable. Using the terms of Fisher's model, Lush demonstrated that only the additive genetic variance contributes to the heritable genetic variance (Lush, 1940). The reason for this is that in diploid outcrossing organisms, each parent contributes one of the two alleles at each gene in an offspring. Thus, the dominance interaction between the two alleles carried by a parent at a locus is not transmitted to its progeny; dominance effects contribute to variation among individuals, but not to the similarity of parent and offspring. So, if the total genotypic variance ($\sigma^2_G$) of a population is composed of additive ($\sigma^2_A$) and dominance ($\sigma^2_D$) variation (and for simplicity ignoring epistatic effects)

($\sigma^2_G = \sigma^2_A + \sigma^2_D$), then the total phenotypic variance is composed of these components plus environmental variance:

$$\sigma^2_P = \sigma^2_G + \sigma^2_\varepsilon = \sigma^2_A + \sigma^2_D + \sigma^2_\varepsilon$$

The proportion of that total phenotypic variance that is heritable is the heritability ($h^2$):

$$h^2 = \frac{\sigma^2_A}{\sigma^2_P} = \frac{\sigma^2_A}{\sigma^2_A + \sigma^2_D + \sigma^2_\varepsilon}$$

Lush (1940) referred to this as the narrow definition of heritability; now commonly referred to as 'heritability in the narrow sense' (Falconer and Mackay, 1996) to distinguish it from a broader definition of heritability ('heritability in the broad sense'):

$$h^2_{BS} = \frac{\sigma^2_G}{\sigma^2_P} = \frac{\sigma^2_A + \sigma^2_D}{\sigma^2_A + \sigma^2_D + \sigma^2_\varepsilon}$$

(Falconer and Mackay, 1996).

These two measures have clear and distinct interpretations. Heritability in the broad sense is the proportion of total phenotypic variation that is because of genotypic differences, and refers to the distinction made by Johannsen between genotypic and phenotypic variation. Heritability in the narrow sense has a very important meaning to breeders because it indicates how effective selection will be; in outcrossing populations, the response to selection ($R$) can be predicted as $R = Sh^2$, where $S$ is the selection differential (the difference between the average of the selected individuals and the whole population). The response to selection refers to the difference between the progeny phenotype values derived from selected parents compared to what the population mean would be expected to be if selection had not occurred (and all parents or a random sample of parents had contributed equally to the next generation).

Lush's definition of narrow-sense heritability is considered the gold standard in quantitative genetics because of its direct relationship with response to selection (Falconer and Mackay, 1996). Note, however, that Lush (an animal breeder) conceived of selection acting on values of individual diploid animals with no possibility of self-fertilization. This is precisely the situation considered by animal breeders and human quantitative geneticists, such that deviations from this definition are considered undesirable and an abuse of the term heritability. Unfortunately, for plant breeders, the reproductive diversity of plants discussed previously immediately causes problems with the definition of heritability (Nyquist, 1991). In addition, polysomic polyploidy of some important crops (e.g., potato, alfalfa, strawberry, many forage, and turf grasses), also creates havoc with the definition of heritability. Finally, the unit of measurement in many plant breeding programs is the total or average production value of a line or family grown in dense populations, such that the measurement of individual values is difficult or meaningless; the result is that the appropriate estimator of heritability in these situations is not the true narrow-sense heritability, but a measure of the heritability of family mean or total values (Falconer and Mackay, 1996; Nyquist, 1991).

## Extending Heritability to Polyploidy and Complex Mating Systems of Plants

To demonstrate the more complicated nature of heritability in plants that can arise even in the case of regular outcrossing similar to the animal breeding situation, consider first a relatively simple case of a tetrasomic tetraploid like potato. Each plant carries four homologous copies of each chromosome type (instead of two copies as in diploids). This means that the genotypic value of a potato plant is the sum of the effects of four alleles at the locus, plus all their possible interactions. Each pair of alleles can have a dominance interaction similar to the case in dipoids, however there is also the possibility of higher-order interactions among the alleles at a single locus – the three- and four-way interactions of the alleles. It would be a helpful simplification to ignore the higher-order interactions; unfortunately, substantial evidence indicates that higher-order allelic interactions have a strong effect on yield and vigor in autotetraploids like potato (Mendoza and Haynes, 1974) and alfalfa (Groose *et al.*, 1989). Even if one could ignore all higher-order allelic interactions in a crop like potato, consider that in regular sexual reproduction, each tetraploid parent contributes two alleles per locus to each offspring. Thus a dominance effect because of the interaction of two alleles can be transmitted from parent to offspring, implying that some proportion of the dominance genetic variation should be admitted as part of the definition of heritability in this system (Holland, 2001; Levings and Dudley, 1963).

The situation becomes more complicated when we introduce the more complex reproductive systems of plants (Nyquist, 1991). Consider further the potato breeding situation outlined above. A potato plant can be intermated sexually to generate outbred progeny, which can be subjected to selection based on phenotype, similar in concept to an animal breeding scheme. The utility of a selected cattle sire in a dairy breeding program, however, is in its ability to be mated to dams to generate daughters that can be used in dairy production. Thus, its breeding value really is determined by its additive genetic effects, because all of its daughters carry only one of its alleles. In contrast, if a potato breeder identifies a particularly superior plant, that plant can be vegetatively propagated by tuber cuttings, and those tuber pieces can be used in production by farmers. Thus, the response to selection in this case involves the whole genotype of the plant that is clonally propagated, including all of its additive and dominance effects. So, the response to this selection is a function of the broad-sense heritability, not the narrow-sense heritability in this case.

At the other extreme, a typical breeding method in self-fertilizing species is to make a cross between two varietal parents, then inbreed the progeny from that mating to nearly complete homozygosity. In a pedigree breeding program, selection may be applied at each generation of this process. Note that the proportion of heterozygosity decreases by half with each generation, so that dominance interactions become progressively less important and more of the additive variation becomes partitioned among inbred lineages than within, until one reaches the situation of genotypic variation partitioned entirely among (and not within) pure lines encountered by Johannsen. Modeling the contribution of additive variation to

the genotypic variation among and within lines at each generation is simply a function of the level of inbreeding. However, modeling the nonadditive components of inheritance that contribute to variation among and within lines under intermediate generations is quite complicated (Cockerham, 1983) and measuring those components is exceedingly difficult (Edwards and Lamkey, 2002). In such a case, the appropriate measure of heritability depends on which generations are considered the 'selection units' and which are considered the 'response units.' In the simplest case, a sample of completely homozygous inbred lines from a cross can be evaluated, and only the best lines selected. If those best lines are then propagated as cultivars and considered the end product of that breeding cycle, then the heritability in this situation is a function of total genotypic variance among completely homozygous lines (which includes twice as much additive variance as in the outbred population, no dominance variance, but possibly other higher-order genetic components of variance). Thus, it is entirely appropriate in this situation to use an estimator of heritability that does not match Lush's narrow-sense heritability. In situations where the selected lines are themselves intermated to form a new breeding population, which itself may be inbred, the appropriate heritability estimator depends on the generation of the selected lines as well as the generation of the progeny lines. For this reason, plant breeders would do well to label clearly their estimates of heritability in terms of the 'units of selection' and the 'units of response' to which they apply (Holland *et al.*, 2003).

## Enhancing Gain from Selection

For a given trait heritability, there is the possibility of enhancing gain from selection by increasing the selection differential ($S$ in the equation above). This can be accomplished by either selecting fewer, more extreme individuals from the population or by growing a larger population sample for a fixed number of selected individuals. However, there is a tradeoff between having to increase resources to evaluate larger population samples versus the detrimental effects of inbreeding and reduced long-term gains from selection that can occur when the number of selected individuals is too small. Most mature plant breeding programs have more or less a fixed amount of evaluation resources and also some target number of selected individuals at each stage, such that the proportion of selected individuals cannot be widely varied without difficulty. An exception to this are some well-capitalized commercial breeding programs for some major crops, such as commercial maize breeding programs, which have increased substantially in scale in recent years (Crosbie *et al.*, 2006) as a means to increase annual gains from selection in a very competitive seed sales market.

Because dramatic changes in the selection differential are often so costly as to be unfeasible, plant breeders have focused intensely on increasing the heritability of important traits as an alternative means to enhance gain from selection. It follows from the definition of heritability that the greater the proportion of additive genetic variance to total phenotypic variance, the greater the expected response from selection. The additive genetic variance is an inherent function of a particular population and generation of inbreeding. Therefore, the breeder has two general ways to increase the heritability of a trait. One is to concentrate breeding on populations that have more genetic variance. The other is to reduce as much as possible the nongenetic contributions to phenotypic variance.

Populations with more genetic variance generally have greater response to selection, so an obvious strategy is to create populations with greater genetic variance. There are two impediments to this. First is the difficulty in identifying those breeding crosses that produce populations with greater genetic variance before they are evaluated, as methods to predict progeny variance based on parental trait or genetic differences have limited reliability (Hung *et al.*, 2012b). Second, breeders usually face a trade-off between the genetic variance of a population and its mean value, because in elite breeding materials, most variation is in the direction of reduced performance. Criteria for selecting among populations include the usefulness statistic that incorporates both the population mean and its variance (Melchinger, 1987; Schnell, 1983) and the mean value of the best families in a population at some consistent percentile (Goodman, 1965), but, again reliable values for these statistics are available only after the populations have been evaluated. When the local elite breeding pool is lacking in sufficient favorable variation, crosses between local and exotic sources of germplasm (perhaps including wild relatives) may be necessary to incorporate the desired genes into a breeding program. Recurrent backcrossing to local materials or longer term planning may be required for such strategies to be successful.

Reduction of environmental contributions to the phenotypic variance is one avenue for improving gain from selection that breeders can often implement without major difficulty. For the purposes of this discussion, it is well to distinguish variation because of macroenvironmental differences (such as the average performance difference of a population among different locations or over years) from variation because of microenvironmental variation that occurs within a particular year and location combination. Further, when breeding populations are evaluated across macroenvironments, the possibility of genotype-by-environment interaction (differences in relative performance of varieties across environments) must be considered (**Figure 2**). A common breeding strategy is to evaluate a set of families across multiple environments and to select based on their average performance. Each family is tested at all environments such that all family means are averaged over all macroenvironment effects, eliminating the macroenvironment effects from the phenotypic variance of family means. (Even if some data are missing from some environments, one can still estimate mean values that take into account the macroenvironment effects so that genotype comparison are not influenced by the macroenvironment effects.) The heritability of family mean differences across environments is

$$h^2 = \frac{\sigma^2_G}{\sigma^2_G + \frac{\sigma^2_{GE}}{e} + \frac{\sigma^2_\varepsilon}{re}}$$

where the genetic variance among families, $\sigma^2_G$, is some function of the additive variance that depends on the type of families

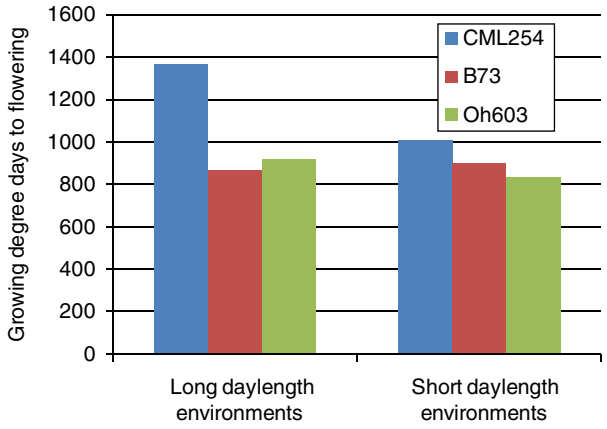


**Figure 2** Example of genotype-by-environment interaction: thermal time to flowering of three maize inbred lines grown in long or short daylength environments. Relative ranking of time to flowering for the three lines changes between these two groups of environments. In this particular case, genetic differences in photoperiod sensitivity among lines determines their reactions to different daylengths.

and their inbreeding level, $\sigma^2_{GE}$ is the genotype-by-environment interaction variance, $\sigma^2_\varepsilon$ is the average within-environment microenvironmental variance, $e$ is the number of environments used for evaluation, and $r$ is the number of replications per family per environment. This equation shows that the heritability, and consequently, the expected gain from selection, can be often increased by increasing the number of testing environments $e$. If the breeder includes environments that are too distinct, however, the genotype-by-environment variance component may increase and overwhelm the genetic variance component, causing heritability to decrease. In such situations, it pays to identify more homogenous subsets of environments and split the breeding program to select for families that are superior within regions, rather than to continue to search for families that are superior across the whole range of macroenvironments (Ceccarelli, 1989; Simmonds, 1991).

Another component of the heritability equation that the breeder can exercise some control over is the within-environment microenvironmental variation (also known as the experimental error variation). As seen in the equation above, the contribution of this component to the denominator of heritability can be reduced by increasing either the number of testing environments or the number of replications at each environment. There is a nearly direct trade-off in resources between increasing the number of replications evaluated and the number of unique genotypes that can be tested. Fortunately, breeders can avail themselves of experimental designs and statistical analysis methods that can help reduce the experimental error variance without requiring additional testing resources. The major challenge to controlling experimental error in breeding trials is the large number of genotypes to be evaluated. With many genotypes to evaluate, the total physical field size required to plant all of them often become so large that the microenvironments of plots can become quite different across the testing field. Most commonly, soil quality may vary across the field so much that comparing the yield of a variety grown on one side of the field to a variety grown on the other side measures more the difference in soil quality than genetic yield potential. Good crop management is critical to minimize such effects, but experimental designs that break the whole field required for a single complete replication of the entries into subblocks can also help significantly. Lattice, alpha, augmented, *p*-rep, and row-column designs are particularly useful (Cullis *et al.*, 2006; Patterson and Williams, 1976; Williams *et al.*, 2011; Wolfinger *et al.*, 1997). In addition to improved experimental design, newer methods that attempt to model the spatial effects on experimental errors can contribute significantly to improving response to selection (Brownie *et al.*, 1993; Gilmour *et al.*, 1997; Qiao *et al.*, 2004; Smith *et al.*, 2005).

Beyond changing the selection differential and heritability, the breeder has an additional avenue through which the response to selection can be improved: reducing the time required to conduct a cycle of selection. Annual crops can be evaluated for only one generation per year in their target production environment. Breeders may be able to use greenhouses or offseason nurseries (in tropical/subtropical locations or in the opposite North/South hemisphere where seasons are offset) to obtain additional generations per year. Unfortunately, greenhouses and offseason nurseries are not useful for evaluating yield potential in the target production environments. Selection aimed at some traits with simple genetic control (such as major gene disease resistances) may be effective, but for the most part, greenhouses and offseason nurseries are most useful for advancing breeding generations (such as selfing, backcrossing, or making new hybrid combinations for breeding or testing) without selection. This alone is useful, but does not directly increase the number of selection cycles per year. Speeding up backcrossing, however, may enable breeders to more quickly introduce new disease resistance genes or transgenes into elite genetic backgrounds, particularly if they are able to easily select for the presence of the desired gene using deoxyribonucleic acid (DNA) tests or simple phenotypes. Lewis and Kernodle (2009) demonstrated an elegant method to dramatically reduce time to flowering in tobacco by transformation with a flowering time promoter, and using these rapid flowering types as targets for backcrossing a desired gene from an undesired genetic background. After the final generation of backcrossing, the breeder can select against the flowering time transgene and for the target gene to recover the desired gene in the genetic background of the elite parent with its normal flowering phenotype. Breeders have searched for other ways to use these additional generations grown outside of their target production environments, and later in this article, a recent approach using DNA marker information alone to conduct selection in these environments will be discussed.

Another way to reduce the time required to conduct a cycle of selection is the use of doubled haploids in species where inbred lines are created as part of the breeding program (Forster and Thomas, 2005). Doubled haploids require special techniques such as anther/pollen culture, wide crossing, embryo rescue, and inducer stocks. The effort required to obtain sufficient numbers of doubled haploid progeny in some cases has limited the utility of doubled haploids. In maize, however, doubled haploid breeding has gone from a rarely used approach to a widespread method in commercial maize

improvement programs in the past decade. This resulted from the development of genetic stocks that combine a relatively high frequency of haploid induction as a pollinator with a simple color marker that permits rapid visual discrimination among diploid, haploid, and outcross progeny to make the recovery of haploids from new breeding crosses relatively simple and efficient (Rober *et al.*, 2005). Proprietary methods to efficiently double the chromosome numbers of large numbers of haploid plants have probably also made an important contribution to this breeding method.

## New Statistical Techniques for Selection

Statistics have long been an important tool for plant breeding; improvements in experimental design have increased the efficiency of selection, as noted in the Section Enhancing Gain from Selection. When experiments are well designed and executed, relatively simple statistical analyses can often remove much of the influence of extraneous macro- and micro-environmental effects from the mean values of the genotypes. To handle nonrandom spatial field effects, however, more complex models may be needed. Such models relaxed the assumptions of standard ordinary least squares analysis of variance, such as the assumption that error effects are uniformly and independently distributed among test plots. Instead, one could test a model that, for example, models the error effects on some pair of plots as having a correlation that is greatest for adjacent plots and decreases in magnitude as the plots considered are more distant physically. To perform such an analysis, however, requires a shift from standard analysis of variance techniques to mixed linear models. The mixed linear models analysis is more computationally intensive and requires specialized software; the necessary computation power and software were not widely available to plant breeders until approximately 20 years ago.

The mixed linear models approach has broadened considerably the range of models that can be tested for plant breeding experiments. One example is the modeling of spatially correlated residual error effects, as just mentioned. More complex models can also be applied to multiple environment trial data sets, such as allowing each environment to have a unique amount of experimental error variance, which is a more realistic model than the constraint of a common error variance assumed by the traditional analysis of variance. An important change in the concept of genotype-by-environment interaction variance has been fostered by the ability to flexibly model different forms of such interactions using mixed models. One example that seems particularly appealing is a model that allows each environment to express a different amount of genotypic variation and allows each pair of environments to have a unique correlation between genotypic values. This allows the breeder to identify easily which pairs of environments produce similar phenotypic responses from a common set of genotypes, and which pairs of environments produce unrelated, or perhaps even negatively correlated phenotypic responses. This helps breeders to select optimal sets of testing environments that will capture as much of the range of environments in the target production region, and can help determine if the breeding program should be partitioned to select for different subsets of genotypes adapted to specific environmental subsets.

Fitting more complex mixed models is not simply an academic exercise in better capturing the nongenetic effects on experiments, however. The mixed model has a different philosophical underpinning for effects that are considered random (meaning that they are samples randomly drawn from some larger population of effects) versus fixed (meaning that the effects studied are the ones of direct interest, and they are not assumed to be drawn from some larger reference population). The practical effect of treating genotypes as random in the mixed model is that their genetic value is predicted with a best linear unbiased predictor (BLUP) instead of estimated with a best linear unbiased estimator (BLUE), such as a mean value (Robinson, 1991). BLUPs and BLUEs may rank genotypes differently, because BLUPs incorporate a shrinkage factor that makes them closer to the overall population mean than their respective BLUEs. The amount of shrinkage depends on the trait heritability (higher heritability traits have less shrinkage) and the amount of information available for a particular genotype. For example, if a genotype has missing data from most environments in a multienvironment trial, but has an outlier phenotype based on a small amount of available data, its BLUP will be shrunk back toward the mean more strongly than genotypes with more data. This is a sort of penalty on those genotypes that have more missing data, and would lead to them being less likely to be selected on the basis of BLUPs, even if their mean BLUE value would have ranked them very highly. The mixed model analysis accounts for the differences among environments in terms of different error variances, amount of genetic variance, correlation with other environments, and so forth, such that the effect of missing data on shrinkage depends on the environment from which the data is missing. Importantly for practical breeding purposes, BLUPs have higher accuracy than mean values, so selection on BLUPs is expected to result in better response to selection (Piepho *et al.*, 2008).

### Best Linear Unbiased Predictions Incorporating Pedigree Information

Another key improvement that mixed models and BLUPs have begun to deliver to plant breeding programs is their ability to incorporate information on relationships among lines and individual plants in the breeding pool. This methodology was pioneered by Henderson (1974) and widely adopted in animal breeding before plant breeders began to recognize its utility and applicability to problems in plant breeding (Bernardo, 1996; Panter and Allen, 1995). Plant breeders long recognized that closely related lines have similar phenotypic responses, but Henderson's BLUP methodology allowed precise incorporation of quantitative estimates of genetic similarity based on the amount of shared pedigree between each pair of lines or individuals into the predictions of genetic value (Lynch and Walsh, 1998). The importance of this method is to leverage observed phenotypic information on each line to improve the predictions of other lines to which they are related (Piepho *et al.*, 2008). Extending this idea further, Bernardo (1996) showed that plant breeders could improve

the efficiency of their programs by initially testing a diverse subset of lines, then predict the genetic value of all of those lines, plus the value of untested lines using the known pedigree relationships among the tested and untested lines and the BLUP method. The lines with highest predicted values (whether directly tested or not) could then be identified for further evaluation, thus concentrating testing resources on lines with the best chances of being superior. This concept has evolved more recently into genomic selection (see Section Using Genetic Markers In Selection Programs), where the concept of breeding value prediction has been further refined through the application of genetic marker information.

## New Sources of Germplasm

A basic premise of plant breeding is that response to selection occurs by the increase of favorable allele frequencies, and concurrent decrease of unfavorable allele frequencies in breeding gene pools. In the long run, plant breeding that is effective should reduce genetic variability in breeding populations. This poses a problem for breeders: progress from selection depends on genetic variation in breeding populations, but such progress tends to ultimately reduce such variation.

Most crops suffered substantial genetic bottlenecks during the domestication process (Tanksley and McCouch, 1997). Domestication tends to impose strong selection for the domestication type: for grain crops, for example, the nonshattering trait greatly simplifies harvest and was strongly selected for across many species (Harlan, 1992). Domestication also imposed significant selection for traits such as loss of seed dormancy and seed size in seed crops, reduction of thorns and awns, and flavor and fruit size in fruit crops (Allard, 1999). In allopolyploid crops such as wheat and oat, multiple bottlenecks occurred: the interspecific hybridizations and subsequent chromosome doubling occurred naturally, but these are very rare events which involve very small population sizes, and were then followed by the typical population bottleneck during domestication (Sears, 1969). This explains the quite low level of genetic variation in wheat compared to other crops (Cox and Wood, 1999). Acting against this domestication bottleneck were subsequent gene flow between sympatric wild and domesticated species (van Heerwaarden *et al.*, 2011) and human selection for very diverse phenotypes (such as the dramatically different plant forms of cabbage noted by Darwin (1859) and for adaptation to widely divergent environments (Allard, 1999; Ruiz *et al.*, 2008; Wallace and Brown, 1988). Mutation, genetic recombination, mating system, and population sizes were also important factors in shaping the diverse levels of genetic variation among crops.

Seed banks and germplasm repositories represent an important genetic resource holding alleles that have been eliminated from modern elite breeding pools but that may have some value to modern agriculture (Tanksley and McCouch, 1997). Major challenges of using gene bank materials are a lack of relevant information on characteristics that may be of use to a particular breeding program and the linkage of favorable alleles to many generally unfavorable alleles at nearby genes in the unadapted genetic backgrounds of most germplasm resources. Harlan and de Wet (1971) proposed a classification of cultivated plants from the point of view of their practical utility to plant breeding programs. In their system, taxa are organized into primary, secondary, and tertiary gene pools. The primary gene pool consists of all taxa that are fully interfertile with the cultivated species; the secondary gene pool includes taxa that can be crossed to the cultivated species but produce hybrids with low fertility or poor chromosome pairing; the tertiary gene pool includes those taxa whose hybrids with the cultivated species are lethal or completely sterile, such that specialized techniques such as embryo rescue and chromosome doubling are required to obtain any fertility in the progeny (Harlan and de Wet, 1971). This classification emphasizes that most breeding crosses should be restricted to the primary gene pool to have the greatest chance of success of recovering an improved cultivar. Nevertheless, even within the primary gene pool of most crops, there are huge differences in adaptation, productivity, acceptability for commercial production, and genetic load. Thus, breeders usually will attempt to incorporate diversity from already improved materials that are distinct from their elite gene pool, but lack significant problems of adaptation or acceptability. Where such materials do not carry the desired favorable alleles, breeders must access less adapted germplasm and expect to need additional generations to break up unfavorable gene linkages and may need to use backcrossing techniques to reduce the overall contribution of the donor of diversity. Despite these difficulties, there is ample evidence that 'exotic' sources of germplasm within the primary gene pool can be effectively used to improve elite gene pools even for highly polygenic traits such as yield (Goodman, 2004; Simmonds, 1993; Tanksley and McCouch, 1997).

For species such as wheat, which suffered severe bottlenecks during their evolution, the primary gene pool may not carry the genetic variation needed for specific traits. Substantial effort in transferring disease and pest resistances from the secondary and tertiary gene pools has been made in wheat, resulting in the development of numerous specialized techniques to recover progeny from wide crosses, induce recombination, and recover stable cultivated types carrying desired target genes (Friebe *et al.*, 1996; Jiang *et al.*, 1993). These approaches, particularly the use of the tertiary gene pool, require longer-term commitment to successfully proceed from initial crosses to an agronomically acceptable cultivar expressing the trait of interest.

In general, the use of secondary and tertiary gene pools is restricted to the introgression of one or a few genes from a particular donor parent. The objective of introgression programs is to recover a variety that carries the target genes from the donor parent but as little genetic material from that donor as possible. This is usually achieved through repeated backcrossing to the recurrent elite cultivated parent, but if the chromosomal structure of the donor is too different from the cultivated parent, the chromosome segments around the target gene may not pair well and recombination may be suppressed. Without adequate recombination, even many generations of backcrossing may result in the maintenance of large chromosomal segments carrying many genes, which, other than the target gene, are likely to be undesirable (Young and Tanksley, 1989).

Introgression programs aimed at introducing genes from the secondary and tertiary gene pool are more likely to be

successful in polyploid species. Successful examples of this work abound for diverse polyploids such as wheat, potato, and sugarcane (Jansky and Peloquin, 2006; Jiang *et al.*, 1993; Simmonds, 1993), but are unheard of in, for example, diploid maize (despite some effort) (Harlan and de Wet, 1977). This is a result of the greater capacity of polyploid genomes to 'buffer' against the deleterious effects of large chromosomal introductions and unbalances that occur during such breeding schemes.

The use of transgenic technologies (gene transformation) can also be considered as part of the toolkit for introducing genes from the tertiary gene pool. Genetic transformation has expanded the tertiary gene pool (at least for some crops) to include all living species, so, for example, the gene that encodes a protein toxic to lepidopteran insects could be transferred from the bacterium *Bacillus thuringiensis* to crop plants (Koziel *et al.*, 1993; Perlak *et al.*, 1990). Genetic transformation is also sometimes considered the ultimate tool for introgression, as only the desired target gene is introduced into the cultivated crop, conceptually eliminating the accompanying linkage drag of the surrounding chromosomal segments from the donor. Unfortunately, in most crops, transformation is most efficiently conducted with older inferior varieties (because they regenerate better from tissue culture), necessitating backcrossing the transgene from the old variety into the elite.

## Using Genetic Markers in Selection Programs

The genomics revolution has provided numerous technologies that facilitate the rapid analysis of sequence variation in crops. Crops such as maize, wheat, rice, and soybean have extensive genomics resources that allow the relatively rapid development of DNA markers that can be used to track inheritance of specific chromosomal positions. Modern sequencing technologies also have contributed to marker development in many other plant species, but the ability to create marker sets that are readily deployable across breeding programs often requires a deeper understanding of genome structure and variation than exists for some species.

The ability to rapidly assay allelic variation at DNA markers can be of considerable utility to plant breeding programs. This technology is of greatest immediate applicability for the selection of desired gene variants that have sizeable effects on an important trait, have been mapped accurately and precisely (with identification of the causal gene and sequence variant most useful), and for which DNA marker assays are cheaper, easier, or faster than phenotypic identification of progenies that carried the desired gene (Collard and Mackill, 2008; Holland, 2004; Tester and Langridge, 2010; Young, 1999). Some examples where this condition is true and for which marker-assisted selection has become routine for at least some breeding programs include:

- Soybean cyst nematode resistance, a difficult trait to accurately phenotype, but for which a few genes have major effects (Cahill and Schmidt, 2004; Young, 1999);
- Disease resistances and submergence tolerance in rice, encompassing a range of difficulty for phenotyping but for which individual genes with large effects have been identified directly or by tight linkage to markers (Collard and Mackill, 2008; Francia *et al.*, 2005; Ismail *et al.*, 2013);
- Numerous disease resistances and simply inherited fruit quality characteristics regularly targeted with markers in commercial tomato breeding programs (Foolad and Panthee, 2012);
- Disease-resistance genes, self-incompatibility, fruit quality in several fruit crops of genus *Prunus* (Dirlewanger *et al.*, 2004);
- Many ($>50$) simply inherited disease resistance and grain quality characteristics targeted by public wheat breeding programs (Dubcovsky, 2004; Eagles *et al.*, 2001; Gupta *et al.*, 2010; Miedaner and Korzun, 2012).

The experience with marker-assisted selection in wheat breeding is particularly instructive, as this has been done almost entirely in the public sector across several countries. Australia pioneered the practical application of marker-assisted selection on a large-scale by creating centralized genotyping facilities to support the applied breeding programs (Eagles *et al.*, 2001). The USA federal government followed suit by establishing four regional small grains genotyping facilities to conduct marker analysis for both public and private sector breeders and by funding a coordinated multistate project to make genotypic selection widely available to applied breeding programs (Dubcovsky, 2004). By removing the burden of marker development and genotyping from each individual breeding program (which typically do not have sufficient resources to perform marker analyses on substantial proportions of their breeding material), the centralized facilities have enabled the integration of marker-assisted selection into cultivar development in small grains. The USA small grains marker-assisted breeding project assisted in the release of at least 90 cultivars.

The wheat example also highlights the trend seen across the examples of marker-assisted selection in other species: specific markers have been most useful for selection when they identify alleles of relatively large effect. Major disease-resistance genes are the most common example of this: the presence of a particular disease-resistance allele can determine completely if a plant is resistant to a particular race of pathogen. In addition to the targeted gene having a major effect, markers are most useful when they are 'diagnostic' for a phenotypic effect across most breeding crosses that are expected to segregate for the phenotype (Collard and Mackill, 2008; Holland, 2004). Although there are special cases in which linkage relationships between linked markers and causal variants tend to hold in breeding programs, high resolution genetic analysis to identify casual sequence variation that underlies desired phenotypes is usually required to obtain a diagnostic marker (Holland, 2004).

Markers have proven particularly useful to backcross breeding programs, where a panel of markers can be used to select simultaneously for the desired allele from the donor parent, against linked donor alleles (to reduce linkage drag), and against donor alleles on other chromosomes (to reduce the number of generations required to recover most of the recurrent parent genome) (Chen *et al.*, 2000; Collard and Mackill, 2008; Frisch *et al.*, 1998; Randhawa *et al.*, 2009).

Implementing markers for selection more generally in typical breeding schemes that involve generating many new breeding families from crosses among numerous parental lines can be more difficult, as it requires diagnostic markers to be widely useful and efficient. Some effort is often required to determine (based on pedigrees and previous knowledge of which founder lines carry specific alleles of interest) which sets of markers can be fruitfully applied for selection in specific breeding populations.

Marker selection can be most useful if it can be applied to single plants (or seeds) in early generations before substantial effort on phenotyping has been expended (Collard and Mackill, 2008). Rather than attempting to identify a plant or plants homozygous for all of the desired marker alleles in early generations (which would require astronomical population sizes if the number of target genes exceeds a few), gene enrichment strategies that select for plants carrying at least one desired allele at as many genes as possible can be far more efficient and practical for implementation in applied programs (Bonnett *et al.*, 2005; Wang *et al.*, 2007). The key concept is to use markers to identify a subset of the breeding population that has the greatest chance of carrying desired alleles for a number of simply inherited traits so that more expensive field phenotyping for other complex traits of agronomic importance can be concentrated on lineages that are most likely to result in a cultivar possessing most of the desired characteristics. Thus, markers do not replace phenotypic evaluations, but instead help to focus phenotyping resources on lines that are enriched for favorable alleles affecting a subset of traits.

In contrast to the situation where trait expression is largely controlled by one or a few genes and for which marker-assisted selection can be very efficient, traits with complex genetic control have been less amenable to marker-assisted selection. Returning again to the example of wheat breeding, an important trait for both yield and grain quality is resistance to Fusarium head blight disease. Unlike some other wheat disease resistances that are largely controlled by a small number of genes with large effects, such as leaf rust and stem rust, Fusarium head blight resistance is under more polygenic control. Numerous genes with smaller effects contribute to the inheritance of this trait and environments exert substantial influence on the expression of this disease. Therefore, it has been difficult to accurately identify the genome locations and effects of the underlying genes. Instead, regions of the genome have been identified as containing quantitative trait loci (QTL) for Fusarium head blight resistance, but these regions are generally only loosely defined and may contain hundreds of genes (Buerstmayr *et al.*, 2009). Worse, the QTL identified in any one linkage mapping study often have little correlation to those identified in a different cross (Buerstmayr *et al.*, 2009). Thus, at this time, breeders have reliably diagnostic markers for only one Fusarium head blight resistance QTL, which confers only a portion of the resistance in some crosses, and has no effect in other crosses (Buerstmayr *et al.*, 2009). The situation is similar for many other complex traits in wheat, such as grain yield and abiotic stress resistances (Francia *et al.*, 2005). There is no evidence for, nor reason to believe, that a particular marker can be diagnostic for any highly polygenic trait in any species (Bernardo, 2008; Collard and Mackill, 2008; Holland, 2004). Statistical difficulties limit the ability to precisely refine the position of QTL or to estimate their effects even in one breeding family (Beavis, 1998; Schön *et al.*, 2004), and the underlying biological reality is that different sets of genes segregate in different breeding families, reducing the predictive power of QTL markers for complex trait selection (Holland, 2007).

Simultaneous genetic analysis of multiple breeding families representing the range of genetic diversity of a breeding pool can more precisely identify specific genes controlling complex traits and at the same time provide better understanding of the distribution of allelic effects among important breeding stocks (Holland, 2004, 2007). Association analysis and joint multiple population linkage analysis represent complementary approaches to comprehensive genetic analysis of complex traits.

Association analysis attempts to identify allelic variants associated with trait variation across a diverse sample of breeding materials. If linkage disequilibrium decays rapidly with physical sequence distance in the sample, association analysis can provide precise identification of the genomic position of sequence variation affecting a trait, possibly facilitating the identification of a causal gene and its causal sequence variation (Ersoz *et al.*, 2007; Flint-Garcia *et al.*, 2005). In addition to potentially higher resolution than linkage mapping, association studies have the benefits of being directly applicable to existing breeding pools without requiring development of specialized mapping families and sampling a greater diversity of alleles (Breseghello and Sorrells, 2006; Ersoz *et al.*, 2007; Myles *et al.*, 2009). A number of experimental and statistical challenges impact the feasibility of association analysis, including the choice of environments used to measure phenotypes of diverse germplasm sampled from a wide range of zones of adaptation, the accuracy of phenotypic measurements, the frequency of the causal allele, and the ability to measure genetic variation for the trait after accounting for population structure (Breseghello and Sorrells, 2006; Larsson *et al.*, 2013; Morrell *et al.*, 2012; Myles *et al.*, 2009; Yan *et al.*, 2011; Zhu *et al.*, 2008). If the genetic regulation of a trait is already well understood from knowledge of biochemistry or model systems, a relatively small number of genes may be considered candidate genes to be tested for association with the trait, increasing one's chance of identifying a causal gene (Harjes *et al.*, 2008; Wilson *et al.*, 2004; Yan *et al.*, 2010). For many important traits, however, one has little or no understanding of the genes or biochemical pathways that might be involved in their expression. In such cases, genome-wide association studies can be attempted, whereby a very large number of single nucleotide polymorphisms (SNPs) are tested for association, in the hope that some causal variants are in linkage disequilibrium with at least one SNP each (Huang *et al.*, 2010, 2012; Morris *et al.*, 2013; Olukolu *et al.*, 2013; Wisser *et al.*, 2011). Genome-wide association studies also have the advantage that they test for variation in noncoding regions of the genome, which turn out to be over represented in the set of SNPs associated with trait variation in some cases (Li *et al.*, 2012).

Joint multiple population linkage analysis combines information from several related biparental cross populations to improve power and resolution of QTL identification and also to characterize the distribution of allele effects in different founder lines representing some portion of the diversity of a

breeding pool (Blanc *et al.*, 2006; Coles *et al.*, 2010; Holland, 2007). The largest such studies in maize have revealed substantial complexity for most quantitative traits, characterized by numerous genes of relatively small effects at which the allelic effects are dispersed among founder lines (Brown *et al.*, 2011; Buckler *et al.*, 2009), demonstrating the futility of predicting QTL effects across populations for highly polygenic traits.

Joint multiple population linkage analysis can be integrated with genome-wide association study to combine the advantages of well-defined population structure and accurate and inexpensive imputation of dense sequence variation from parents to mapping progenies from linkage analysis to the higher resolution provided by association analysis (Hung *et al.*, 2012a; Kump *et al.*, 2011; Tian *et al.*, 2011). These approaches have permitted the identification of SNPs and other sequence variants associated with complex traits at high power, but again reveal the substantial complexity of quantitative traits at least in outcrossing maize, because most SNPs are associated with only a very small fraction of the observed trait variation. Thus, even identification of specific sequence variants associated with trait variation and their distribution in the breeding pool is not sufficient to enable cost-effective marker-assisted selection, because many markers need to be assayed and the response to selection at any one marker is expected to be too small to warrant its selection. However, even for complex traits, there is usually a distribution of variant effect sizes, and perhaps a few variants might have effects of sufficient magnitude to warrant their targeting for marker selection (Hung *et al.*, 2012a). In addition, association analysis might still be fruitfully applied to traits that are difficult to phenotype but for which a relatively small number of genes are important (Harjes *et al.*, 2008).

An entirely distinct approach to the use of DNA markers for enhancing selection to polygenic traits referred to as genomic selection has developed rapidly in the past decade as a way to address the shortcomings of QTL-based marker-assisted selection. Once again, the seminal work in this area of quantitative genetics-based breeding was done by animal breeders (Hayes *et al.*, 2009; Meuwissen *et al.*, 2001). Plant breeders realized the potential utility of genomic selection and proposed modified schemes that could be integrated into current maize (Bernardo and Yu, 2007), small grains (Heffner *et al.*, 2009; Lorenz *et al.*, 2011), and forest tree (Grattapaglia and Resende, 2011) breeding schemes.

As mentioned already, QTL effects for polygenic traits are difficult to estimate accurately, and this limits the predictive ability of QTL-statistical models even within a single mapping family (Melchinger *et al.*, 1998; Schön *et al.*, 2004), and thus the potential response to marker-assisted selection of QTL (Moreau *et al.*, 1998). QTL models are created by testing each region of the genome for association with the trait and selecting only those marker-tagged regions that appear to have statistically significant effects on the trait. Although the exclusion of markers that do not have strong statistical significance from the prediction model seems sensible, it turns out to limit the predictive power of the models (Moreau *et al.*, 1998). The basic premise of genomic selection is that all markers should be included in the prediction model (Bernardo and Yu, 2007; Meuwissen *et al.*, 2001; Xu, 2003), even when there are more markers than individuals in the test population! Standard regression techniques cannot be used for such 'overfitted' models, so alternative statistical techniques must be used, such as Bayesian analysis, ridge regression, or mixed models with constrained marker variances (Bernardo and Yu, 2007; Meuwissen *et al.*, 2001; Xu, 2003). The details of these analyses are complex, but the key objective of genomic selection is to predict the breeding value of lines in the study as well as lines with genotype information but no phenotype information using the prediction model (Lorenz *et al.*, 2011). Unlike QTL mapping, the prediction models do not attempt to accurately estimate the effects of each genome region, instead they sacrifice accuracy on individual marker effects by overfitting the markers, but increase the accuracy of breeding value prediction based on the net value of markers. In this way, the optimal amount of information about polygenic effects is extracted from the set of lines with both genotypic and phenotypic information (the training data set), providing better predictions for lines with only genotypic data (Heffner *et al.*, 2009), which could represent untested sib lines or progeny generated from crosses among lines in the training data set. Genomic selection models are not expected to provide accurate estimates of the effects of specific genome regions on traits, however.

Bernardo and Yu's (2007) original proposal for implementing genomic selection was in the context of a commercial maize breeding program, in which genomic recurrent selection could be conducted for several cycles on seeds or seedlings in offseason nurseries. Separate genomic selection models could be created for each biparental cross family, which maximizes the consistency of linkage relationships between markers and causal genes over several cycles of selection. Even when genomic marker predictions of phenotypes are less accurate than direct phenotypic evaluations, genomic selection can produce greater gains per unit of time by enabling additional cycles of selection on individuals seeds or plants in offseason nurseries (Heffner *et al.*, 2011a, 2010; Lorenzana and Bernardo, 2009; Massman *et al.*, 2013).

An alternative use of genomic selection is the development of prediction models encompassing lines derived from many different parental combinations rather than just one cross (Crossa *et al.*, 2010; Heffner *et al.*, 2010; Zhong *et al.*, 2009). Cross-validation and simulation studies suggest that this could work (Crossa *et al.*, 2010; Heffner *et al.*, 2011b; Riedelsheimer *et al.*, 2012), but that there are difficulties with combining information across very diverse germplasm sets (Lorenz *et al.*, 2012; Zhong *et al.*, 2009). Windhausen *et al.* (2012) demonstrated that genomic prediction models created by combining diverse maize breeding pools had good accuracy for prediction of germplasm in those same pools, but the prediction accuracy fell to approximately zero when the models were applied to newly created biparental populations. Simply, the training data set used to create the genomic selection model must have a close genetic relationship to the breeding population to which it is derived. Exactly how close this relationship must be is still a matter of investigation. The availability of such large-scale prediction models is likely most useful for well-resourced commercial breeding programs, where the ability to generate and genotype new progeny lines outstrips the ability to conduct high-quality phenotypic evaluations of yield and other complex traits. In this way, the breeding potential of progeny

lines that have never been planted in a field could be predicted, and lines with superior predicted values could be retrieved from storage for future phenotypic evaluations. This approach emphasizes the expenditure of precious phenotypic testing resources for materials that have the best chance of being cultivars or cultivar parents.

Marker-assisted selection and genomic selection are important components of modern breeding programs in many crops. The balance between the use of resources on DNA markers to select for specific gene alleles, genomic selection to select for the 'polygenic background,' and phenotypic evaluations will likely shift as research indicates the best application of each evaluation method. The optimal balance likely will vary among crops and even among different programs in the same crop, as it will depend on the relative availability and cost of genomics resources compared to field testing.

To the extent that genetic dissections of important traits succeed in identifying causal genes and sequence variants (via association analysis, high resolution linkage mapping, or other means), those alleles can be targeted for selection and for incorporation across distinct germplasm groups. This form of 'direct allele selection' (Sorrells and Wilson, 1997) should be more effective at predicting the value of alleles across diverse germplasm than genomic selection models, which are highly dependent on the genetic context in which they are defined. These different selection methods could be combined by identifying a subset of the most important genes and targeting them for direct allele selection, predicting the breeding value of some portion of the remaining background polygenic variation with a genomic selection model, and relying on extensive phenotypic evaluations of a selected subset of lines to make final decisions on cultivar releases. Note that high-quality phenotypic evaluations underlie all three of these aspects of modern breeding.

*See also*: Crop Pollination. Genebanks: Past, Present, and Optimistic Future. Genomics: Plant Genetic Improvement. Green Revolution: Past, Present, and Future. Transgenic Methodologies – Plants

## References


Allard, R.W., 1960. Principles of Plant Breeding. New York: Wiley.

Allard, R.W., 1999. Principles of Plant Breeding, second ed. New York: John Wiley and Sons.

Beavis, W.D., 1998. QTL analyses: Power, precision, and accuracy. In: Paterson, A. H. (Ed.), Molecular Dissection of Complex Traits. Boca Raton, FL: CRC Press, pp. 145–162.

Bernardo, R., 1996. Best linear unbiased prediction of maize single-cross performance. Crop Science 36, 50–56.

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

Bernardo, R., Yu, J., 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Science 47, 1082–1090.

Blanc, G., Charcosset, A., Mangin, B., Gallais, A., Moreau, L., 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: An application in maize. Theoretical and Applied Genetics 113, 206–224.

Bonnett, D.G., Rebetzke, G.J., Spielmeyer, W., 2005. Strategies for efficient implementation of molecular markers in wheat breeding. Molecular Breeding 15, 75–85.

Breseghello, F., Sorrells, M.E., 2006. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science 46, 1323–1330.

Brown, P.J., Upadyayula, N., Mahone, G., *et al*., 2011. Distinct genetic architectures for male and female inflorescence traits of maize. PLoS Genetics 7, e1002383.

Brownie, C., Bowman, D.T., Burton, J.W., 1993. Estimating spatial variation in analysis of data from yield trials: A comparison of methods. Agronomy Journal 85, 1244–1253.

Buckler, E.S., Holland, J.B., McMullen, M.M., *et al*., 2009. The genetic architecture of maize flowering time. Science 325, 714–718.

Buerstmayr, H., Ban, T., Anderson, J.A., 2009. QTL mapping and marker-assisted selection for fusarium head blight resistance in wheat: A review. Plant Breeding 128, 1–26.

Cahill, D.J., Schmidt, D.H., 2004. Use of marker assisted selection in a product development breeding program. In Fischer, T., Turner, N., Angus, J., *et al*., (Eds.), New Directions for a Diverse Planet: Proceedings of 4th International Crop Science Congress. Gosford, NSW, Australia: The Regional Institute Ltd.

Ceccarelli, S., 1989. Wide adaptation: How wide? Euphytica 40, 197–205.

Chen, S., Lin, X.H., Xu, C.G., Zhang, Q., 2000. Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Science 40, 239–244.

Citrus Breeding 105–134.

Cockerham, C.C., 1983. Covariances of relatives from self-fertilization. Crop Science 23, 1177–1180.

Coles, N.D., McMullen, M.D., Balint-Kurti, P.J., Pratt, R.C., Holland, J.B., 2010. Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. Genetics 184, 799–812.

Collard, B.C.Y., Mackill, D.J., 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 557–572.

Cox, T.S., Wood, D., 1999. The nature and role of crop biodiversity. In: Wood, D., Lenne, J.M. (Eds.), Agrobiodiversity: Characterization, Utilization, and Management. Wallingford, UK: CABI Publishing, pp. 35–37.

Crosbie, T.M., Eathington, S.R., Johnson, G.R., *et al*., 2006. Plant breeding: Past, present, and future. In: Lamkey, K.R., Lee, M. (Eds.), Plant breeding: The Arnel R. Hallauer International Symposium. Ames, IA: Blackwell. pp. 3–50.

Crossa, J., de los Campos, G., Perez, P., *et al*., 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186, 713–724.

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

Darwin, C., 1859. On the Origin of Species. London: John Murrawy.

Dirlewanger, E., Graziano, E., Joobeur, T., *et al*., 2004. Comparative mapping and marker-assisted selection in rosaceae fruit crops. Proceedings of the National Academy of Sciences of the USA 101, 9891–9896.

Dubcovsky, J., 2004. Marker-assisted selection in public breeding programs: The wheat experience. Crop Science 44, 1895–1898.

Duvick, D.N., Smith, J.S.C., Cooper, M., 2004. Changes in performance, parentage, and genetic diversity of successful corn hybrids, 1930–2000. In: Smith, C.W., Betran, F.J., Runge, E.C.A. (Eds.), Corn: Origin, History, Technology, and Production. New York: Wiley, pp. 65–97.

Eagles, H.A., Bariana, H.S., Ogbonnaya, F.C., *et al*., 2001. Implementation of markers in Australian wheat breeding. Australian Journal of Agricultural Research 52, 1349–1356.

East, E.M., 1910. A Mendelian interpretation of variation that is apparently continuous. American Naturalist 44, 65–82.

East, E.M., 1916. Studies on size inheritance in Nicotiana. Genetics 1, 161–176.

Edwards, J.E., Lamkey, K.R., 2002. Quantitative genetics of inbreeding in a synthetic maize population. Crop Science 42, 1094–1104.

Ersoz, E.S., Yu, J., Buckler, E.S., 2007. Applications of linkage disequilibrium and association mapping in crop plants. In: Varshney, R.K., Tuberosa, R. (Eds.), Genomics-Assisted Crop Improvement, vol. 1. New York: Springer, pp. 97–119.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to quantitative genetics, fourth ed. Essex, UK: Longman Technical.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society, Edinburgh 52, 399–433.

Flint-Garcia, S.A., Thuillet, A.C., Yu, J., *et al*., 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant Journal 44, 1054–1064.

Foolad, M.R., Panthee, D.R., 2012. Marker-assisted selection in tomato breeding. Critical Reviews in Plant Sciences 31, 93–123.

Forster, B.P., Thomas, W.T., 2005. Doubled haploids in genetics and plant breeding. Plant Breeding Reviews 25, 57–88.

Francia, E., Tacconi, G., Crosatti, C., *et al.*, 2005. Marker assisted selection in crop plants. Plant Cell, Tissue and Organ Culture 82, 317–342.

Friebe, B., Jiang, J., Raupp, W., McIntosh, R., Gill, B., 1996. Characterization of wheat-alien translocations conferring resistance to diseases and pests: Current status. Euphytica 91, 59–87.

Frisch, M., Bohn, M., Melchinger, A.E., 1998. Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Science 39, 1295–1301.

Gilmour, A.R., Cullis, B.R., Verbyla, A.P., 1997. Accounting for natural and extraneous variation in the analysis of field experiments. Journal of Agricultural, Biological, and Environmental Statistics 2, 269–293.

Gmitter, F.G., Soneji, J.R., Rao, M.N., 2009. Citrus breeding. In: Gradziel, T.M. (Ed.), Breeding Plantation Tree Crops: Temperate Species. Springer: New York, NY, pp. 105–134.

Goodman, M.M., 1965. Estimates of genetic variance in adapted and exotic populations of maize. Crop Science 5, 87–90.

Goodman, M.M., 2004. Developing temperate inbreds using tropical maize germplasm: Rationale, results, conclusions. Maydica 49, 209–219.

Grattapaglia, D., Resende, M.D., 2011. Genomic selection in forest tree breeding. Tree Genetics and Genomes 7, 241–255.

Groose, R.W., Talbert, L.E., Kojis, W.P., Bingham, E.T., 1989. Progressive heterosis in autotetraploid alfalfa: Studies using two types of inbreds. Crop Science 29, 1173–1177.

Gupta, P., Langridge, P., Mir, R., 2010. Marker-assisted wheat breeding: Present status and future possibilities. Molecular Breeding 26, 145–161.

Harjes, C.E., Rocheford, T.R., Bai, L., *et al.*, 2008. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319, 330–333.

Harlan, J.R., 1992. Crops and Man, second ed. Madison, WI: American Society of Agronomy.

Harlan, J.R., de Wet, J.M.J., 1971. Toward a rational classification of cultivated plants. Taxon 20, 509–517.

Harlan, J.R., de Wet, J.M.J., 1977. Pathways of genetic transfer from *Tripsacum* to *Zea mays*. Proceedings of the National Academy of Sciences of the USA 74, 3494–3497.

Hayes, B., Bowman, P., Chamberlain, A., Goddard, M., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science 92, 433.

van Heerwaarden, J., Doebley, J., Briggs, W.H., *et al.*, 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proceedings of the National Academy if Sciences of the USA 108, 1088–1092.

van Heerwaarden, J., Hufford, M.B., Ross-Ibarra, J., 2012. Historical genomics of north american maize. Proceedings of the National Academy of Sciences of the USA 109, 12420–12425.

Heffner, E.L., Jannink, J.-L., Iwata, H., Souza, E., Sorrells, M.E., 2011a. Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Science 51, 2597–2606.

Heffner, E.L., Jannink, J.-L., Sorrells, M.E., 2011b. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4, 65–75.

Heffner, E.L., Lorenz, A.J., Jannink, J.L., Sorrells, M.E., 2010. Plant breeding with genomic selection: Gain per unit time and cost. Crop Science 50, 1681–1690.

Heffner, E.L., Sorrells, M.E., Jannink, J.L., 2009. Genomic selection for crop improvement. Crop Science 49, 1–12.

Henderson, C.R., 1974. General flexibility of linear model techniques for sire evaluation. Journal of Dairy Science 57, 963–972.

Holland, J.B., 2001. Epistasis and plant breeding. Plant Breeding Reviews 21, 27–92.

Holland, J. B., 2004. Implementation of molecular markers for quantitative traits in breeding programs – challenges and opportunities. In Fischer, T., Turner, N., Angus, J., *et al.* (Eds.), New Directions for a Diverse Planet: Proceedings of 4th International Crop Science Congress. Gosford, NSW, Australia: The Regional Institute Ltd.

Holland, J.B., 2007. Genetic architecture of complex traits in plants. Current Opinion in Plant Biology 10, 156–161.

Holland, J.B., Nyquist, W.E., Cervantes-Martinez, C.T., 2003. Estimating and interpreting heritability for plant breeding: An update. Plant Breeding Reviews 22, 9–111.

Huang, X., Wei, X., Sang, T., *et al.*, 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Genetics 42, 961–967.

Huang, X., Zhao, Y., Wei, X., *et al.*, 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature Genetics 44, 32–39.

Hung, H.-Y., Shannon, L.M., Tian, F., *et al.*, 2012a. *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proceedings of the National Academy of Sciences of the USA 109, E1913–E1921.

Hung, H.Y., Browne, C., Guill, K., *et al.*, 2012b. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. Heredity 108, 490–499.

Ismail, A.M., Singh, U.S., Singh, S., Dar, M.H., Mackill, D.J., 2013. The contribution of submergence-tolerant (*sub1*) rice varieties to food security i n flood-prone rainfed lowland areas in Asia. Field Crops Research 152, 83–93.

Janik, J., 2006. Origins of fruit culture and fruit breeding. In: Lamkey, K.R., Lee, M. (Eds.), Plant breeding: The Arnel. R. Hallauer International Symposium. Ames, IA: Blackwell. pp. 269–282.

Jansky, S.H., Peloquin, S.J., 2006. Advantages of wild diploid *Solanum* species over cultivated diploid relatives in potato breeding programs. Genetic Resources and Crop Evolution 53, 669–674.

Jiang, J., Friebe, B., Gill, B.S., 1993. Recent advances in alien gene transfer in wheat. Euphytica 73, 199–212.

Johannsen, W., 1903. Ueber erblichkeit in populationen und in reinen leinen. Jena: Gustav Fischer.

Johannsen, W., 1911. The genotype conception of heredity. American Naturalist 45, 129–159.

Koziel, M.G., Beland, G.L., Bowman, C., *et al.*, 1993. Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. Biotechnology 11, 194–200.

Kump, K.L., Bradbury, P.J., Buckler, E.S., *et al.*, 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nature Genetics 43, 163–169.

Larsson, S.J., Lipka, A.E., Buckler, E.S., 2013. Lessons from *dwarf8* on the strengths and weaknesses of structured association mapping. PLoS Genetics 9, e1003246.

Levings, C.S., Dudley, J.W., 1963. Evaluation of certain mating designs for estimation of genetic variance in autotetraploid alfalfa. Crop Science 3, 532–535.

Lewis, R.S., Kernodle, S.P., 2009. A method for accelerated trait conversion in plant breeding. Theoretical and Applied Genetics 118, 1499–1508.

Li, J., Yuan, L., 2010. Hybrid Rice: Genetics, Breeding, and Seed Production Plant Breeding Reviews. Oxford, UK: John Wiley & Sons, Inc. 15–158.

Li, X., Zhu, C., Yeh, C.-T., *et al.*, 2012. Genic and nongenic contributions to natural variation of quantitative traits in maize. Genome Research 22, 2436–2444.

Lorenz, A.J., Chao, S.M., Asoro, F.G., *et al.*, 2011. Genomic selection in plant breeding: Knowledge and prospects. Advances in Agronomy 110, 77–123.

Lorenz, A.J., Smith, K.P., Jannink, J.-L., 2012. Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. Crop Science 52, 1609–1621.

Lorenzana, R.E., Bernardo, R., 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theoretical and Applied Genetics 120, 151–161.

Lush, J.L., 1940. Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. Journal of Animal Science 33, 293–301.

Lush, J.L., 1945. Animal Breeding Plans, third ed. Ames, IA: Collegiate Press.

Lynch, M., Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sunderland, MA: Sinauer Associates, Inc.

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.

Melchinger, A.E., 1987. Expectation of means and variances of testcrosses produced from $F_2$ and backcross individuals and their selfed progenies. Heredity 59, 105–115.

Melchinger, A.E., Utz, H.F., Schön, C.C., 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveal low power of QTL detection and large bias in estimates of QTL effects. Genetics 149, 383–403.

Mendoza, H., Haynes, F., 1974. Genetic basis of heterosis for yield in the autotetraploid potato. Theoretical and Applied Genetics 45, 21–25.

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

Miedaner, T., Korzun, V., 2012. Marker-assisted selection for disease resistance in wheat and barley breeding. Phytopathology 102, 560–566.

Moreau, L., Charcosset, A., Hospital, F., Gallais, A., 1998. Marker-assisted selection efficiency in populations of finite size. Genetics 148, 1353–1365.
Morrell, P.L., Buckler, E.S., Ross-Ibarra, J., 2012. Crop genomics: Advances and applications. Nature Review Genetics 13, 85–96.
Morris, G.P., Ramu, P., Deshpande, S.P., *et al.*, 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proceedings of the National Academy of Sciences of the USA 110, 453–458.
Myles, S., Peiffer, J., Brown, P.J., *et al.*, 2009. Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell 21, 2194–2202.
Nyquist, W.E., 1991. Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Science 10, 235–322.
Olukolu, B.A., Negeri, A., Dhawan, R., *et al.*, 2013. A connected set of genes associated with programmed cell death implicated in controlling the hypersensitive response in maize. Genetics 193, 609–620.
Ortiz, R., Dochez, C., Asiedu, R., Moonan, F., 2006. Breeding vegetatively propagated crops. In: Lamkey, K.R., Lee, M. (Eds.), Plant breeding: The Arnel Hallauer International Symposium. Ames, IA: Blackwell, pp. 251–268.
Panter, D.M., Allen, F.L., 1995. Using best linear unbiased predictions to enhance breeding for yield in soybean: I. Choosing parents. Crop Science 35, 397–405.
Patterson, H.D., Williams, E.R., 1976. A new class of resolvable incomplete block designs. Biometrika 63, 83–92.
Perlak, F.J., Deaton, R.W., Armstrong, T.A., *et al.*, 1990. Insect resistant cotton plants. Biotechnology 8, 939–943.
Piepho, H.-P., Mohring, J., Melchinger, A.E., Buchse, A., 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161, 209–228.
Qiao, C.G., Basford, K.E., DeLacy, I.H., Cooper, M., 2004. Advantage of single-trial models for response to selection in wheat breeding multi-environment trials. Theoretical and Applied Genetics 108, 1256–1264.
Randhawa, H.S., Mutti, J.S., Kidwell, K., *et al.*, 2009. Rapid and targeted introgression of genes into popular wheat cultivars using marker-assisted background selection. PLoS One 4, e5752.
Richards, A.J., 1997. Plant Breeding Systems, second ed. UK: Chapman & Hall.
Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., *et al.*, 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics 44, 217–220.
Rober, F., Gordillo, G., Geiger, H., 2005. *In vivo* haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica 50, 275.
Robinson, G.K., 1991. That BLUP is a good thing: The estimation of random effects. Statistical Science 6, 15–51.
Ruiz Corral, J.A., Puga, N.D., Sánchez Gonzalez, J.D.J., *et al.*, 2008. Climatic adaptation and ecological descriptors of 42 Mexican maize races. Crop Science 48, 1502–1512.
Schnell, F.W., 1983. Probleme der elternwahl-ein überblick. Arbeitstagung der Arbeitsgemeinschaft der Saatzuchleiter in Gumpenstein 1–11.
Schön, C.C., Utz, H.F., Groh, S., *et al.*, 2004. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167, 485–498.
Sears, E.R., 1969. Wheat cytogenetics. Annual Review of Genetics 3, 451–468.
Simmonds, N.W., 1991. Selection for local adaptation in a plant breeding programme. Theoretical and Applied Genetics 82, 363–367.
Simmonds, N.W., 1993. Introgression and incorporation. Strategies for the use of crop genetic resources. Biological Reviews 68, 539–562.
Smith, A.B., Cullis, B.R., Thompson, R., 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. Journal of Agricultural Science 143, 1–14.
Sorrells, M.E., Wilson, W.A., 1997. Direct classification and selection of superior alleles for crop improvement. Crop Science 37, 691–697.
Stebbins, G.L., 1974. Flowering Plants: Evolution Above the Species Level. Cambridge, MA: Belknap Press.
Tanksley, S.D., McCouch, S.R., 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277, 1063–1066.
Tester, M., Langridge, P., 2010. Breeding technologies to increase crop production in a changing world. Science 327, 818–822.
Tian, F., Bradbury, P.J., Brown, P.J., *et al.*, 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics 43, 159–162.
Wallace, H.A., Brown, W.L., 1988. Corn and Its Early Fathers. Ames, IA: Iowa State University Press Rev.
Wang, J., Chapman, S.C., Bonnett, D.G., Rebetzke, G.J., Crouch, J., 2007. Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Science 47, 582–588.
Williams, E., Piepho, H.-P., Whitaker, D., 2011. Augmented p-rep designs. Biometrical Journal 53, 19–27.
Wilson, J.A., 1984. Hybrid Wheat Breeding and Commercial Seed Development Plant Breeding Reviews. Hoboken, NJ, USA: John Wiley & Sons, Inc. 303–319.
Wilson, L.M., Whitt, S.R., Ibañez, A.M., *et al.*, 2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16, 2719–2733.
Windhausen, V.S., Atlin, G.N., Hickey, J.M., *et al.*, 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3: Genes: Genomes: Genetics 2, 1427–1436.
Wisser, R.J., Kolkman, J.M., Patzoldt, M.E., *et al.*, 2011. Multivariate analysis of maize disease resistances suggests a pleiotropic genetic basis and implicates a *GST* gene. Proceedings of the National Academy of Sciences of the USA 108, 7339–7344.
Wolfinger, R.D., Federer, W.T., Corderobrana, O., 1997. Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. Agronomy Journal 89, 856–859.
Wu, K.-K., Ming, R., Moore, P.H., Paterson, A.H., 2006. Sugarcane genomics and breeding. In: Lamkey, K.R., Lee, M. (Eds.), Plant Breeding: The Arnel R. Hallauer International Symposium. USA: Blackwell Publishing, pp. 283–292.
Xu, S., 2003. Estimating polygenic effects using markers of the entire genome. Genetics 163, 789–801.
Yan, J., Warburton, M., Crouch, J., 2011. Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Science 51, 433–449.
Yan, J.B., Kandianis, C.B., Harjes, C.E., *et al.*, 2010. Rare genetic variation at *Zea mays crtrb1* increases beta-carotene in maize grain. Nature Genetics 42, 322–327.
Young, N.D., 1999. A cautiously optimistic vision for marker-assisted breeding. Molecular Breeding 5, 505–510.
Young, N.D., Tanksley, S.D., 1989. RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theoretical and Applied Genetics 77, 353–359.
Zhong, S., Dekkers, J.C.M., Fernando, R.L., Jannink, J.-L., 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. Genetics 182, 355–364.
Zhu, C., Gore, M., Buckler, E.S., Yu, J., 2008. Status and prospects of association mapping in plants. The Plant Genome 1, 5–20.

## Relevant Websites

http://www.cgiar.org/
Consultative Group on International Agricultural Research, which organizes research conducted at non-profit international research centers.

http://www.ars-grin.gov/
Germplasm collections of USDA, including searchable databases of seed collections.

http://maswheat.ucdavis.edu
MASwheat collaborative project on marker-assisted selection in wheat, including technical information on laboratory protocols and outreach and educational materials for the general public.

http://www.extension.org/plant_breeding_genomics
On-line course materials and short tutorials on specific topics from extension Foundation, a network of agricultural extension services of USA land-grant universities.

http://www.plantbreeding.org/napb/
USA National Association of Plant Breeders, includes the plant breeding newsletter.