Annual Review of Genetics

On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics

Jason G. Wallace,[1] Eli Rodgers-Melnick,[2] and Edward S. Buckler[3,4]

[1]Department of Crop and Soil Sciences, The University of Georgia, Athens, Georgia 30602, USA; email: jason.wallace@uga.edu

[2]Corteva Agriscience, DowDuPont, Johnston, Iowa 50131, USA

[3]United States Department of Agriculture, Agricultural Research Service, Ithaca, New York 14853, USA

[4]Institute for Genomic Diversity, Cornell University, Ithaca, New York 14853, USA

Annu. Rev. Genet. 2018. 52:26.1–26.24

The *Annual Review of Genetics* is online at genet.annualreviews.org

https://doi.org/10.1146/annurev-genet-120116-024846

Keywords

quantitative genetics, breeding, agriculture, adaptation, heterosis, deleterious alleles

Abstract

Understanding the quantitative genetics of crops has been and will continue to be central to maintaining and improving global food security. We outline four stages that plant breeding either has already achieved or will probably soon achieve. Top-of-the-line breeding programs are currently in Breeding 3.0, where inexpensive, genome-wide data coupled with powerful algorithms allow us to start breeding on predicted instead of measured phenotypes. We focus on three major questions that must be answered to move from current Breeding 3.0 practices to Breeding 4.0: (*a*) How do we adapt crops to better fit agricultural environments? (*b*) What is the nature of the diversity upon which breeding can act? (*c*) How do we deal with deleterious variants? Answering these questions and then translating them to actual gains for farmers will be a significant part of achieving global food security in the twenty-first century.

INTRODUCTION

Quantitative genetics: the study of traits that vary continuously in a population instead of having only a few discrete states

The outlook for future food security is well known: By 2050 there will be approximately 9 billion people on the planet, and 9–11 billion by 2100 (50). The linear growth in food production seen historically will not be enough to satisfy global demand by 2050 (142), especially with increasing demand for high-quality protein by a growing global middle class (171). This argument is as old as Malthus, yet to date agricultural production has managed to keep ahead of population growth thanks to mechanization, fertilization, plant breeding, and other agronomic advances. However, it is dangerous to assume that just because production has outpaced population growth in the past that it will inevitably do so in the future. Most of the easily arable land has already been brought into cultivation, and owing to land degradation and urbanization, its area is actually shrinking (198). Most major sources of water—both surface and groundwater—are being overdrawn, and water shortages are likely to get worse in the next few decades (54). In addition, extreme weather events, such as droughts, floods, and damaging storms, are predicted to increase as climate change moves through the twenty-first century (17, 25, 27). In short, it is getting harder and harder to grow crops, yet more and more people rely on us doing so.

THE FOUR STAGES OF PLANT BREEDING

Only 1–3% of modern industrialized society is directly involved in food production (187), a sharp decline from rates in historical societies. Much of this shift is the result of management improvements (e.g., plows, planters, fertilizers), but much is also due to improvements in crop genetics through breeding (37). Although some important agricultural traits have discrete genetics–such as Mendel's famous pea phenotypes or major disease resistance loci–most are highly quantitative, such as plant architecture, maturity, nutritional quality, and yield. Although quantitative genetics is roughly a century old (43, 47, 188), the principles it encompasses have been applied throughout the history of plant breeding.

Based on the techniques involved, we have split plant breeding into four major stages: three that have already been achieved and one that we will likely reach in the near future (**Figure 1**). Each of these stages builds upon earlier ones by integrating established techniques with new technologies to increase breeding efficiency.

Breeding 1.0 began 10–12 thousand years ago, as people from around the world explored and cultivated nearly 7,000 species of food plants (81). Professional breeders were rare or absent, but phenotype-based selection by local, independent farmers eventually resulted in the dramatic phenotypic changes seen in modern crops (36, 120).

Breeding 2.0 began in the late nineteenth and early twentieth centuries when inbreeding depression was recognized (32), Mendelian genetics was (re)discovered (26, 34, 173), and quantitative genetics theory was established (43, 188). Many advances in plant breeding during this time were in the science of breeding itself, including replicated field trials, controlled crossings, statistical analyses, formal experimental designs, hybrid breeding, pedigree-based estimates of breeding values, and precise measurement of yield at scale (e.g., with multirow combines).

Roughly 30 years ago we entered Breeding 3.0 when molecular markers and genomic data began to complement phenotypic data. This stage started with marker-assisted backcrossing and pedigree confirmation, then moved to dissecting complex traits with linkage mapping (91). The introduction of high-throughput genotyping then expanded the quantitative genetics tool kit to dissect variation in natural populations (genome-wide association) (147) and to select on genome-estimated breeding values (genomic selection) (119).

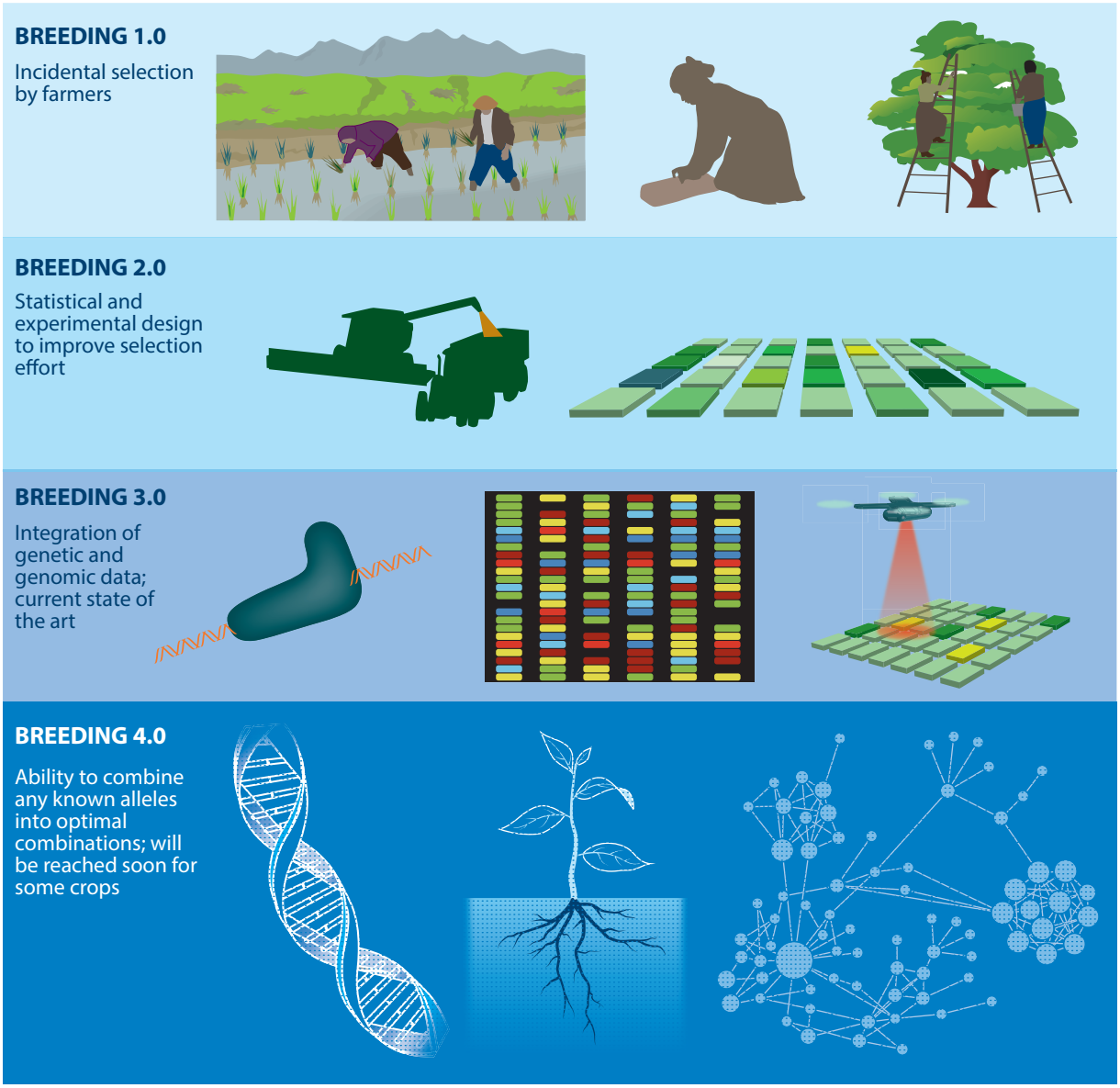

Figure 1

Four stages of plant breeding. Breeding efforts can be divided into four existing or near-future stages based on available methodology. Breeding 1.0 is mostly incidental selection by farmers. Breeding 2.0 involves using statistics and experimental design to improve selection efforts. Breeding 3.0 includes genetic and genomic data and is the current state of the art. Breeding 4.0 will probably arrive soon (at least for some crops) and is marked by the genome-wide ability to combine any known alleles into desirable combinations.

We are now on the cusp of Breeding 4.0, a new level of breeding where functional genetic variants can be rationally combined both faster and better than ever before. This level of breeding is catalyzed by major technological advances in genetics and information systems. For example, genome resequencing studies can now cost less than a replicated yield trial, and genome editing is

expected to enable parallel, precise modifications to many sites (possibly hundreds) per generation. High-throughput phenotyping can measure numerous traits with unprecedented spatiotemporal resolution, and machine learning approaches permit the processing and interpretation of agronomic data at a level far beyond what humans can assimilate.

Although one can conceive an even more advanced level (Breeding 5.0) involving de novo design of genes, pathways, and traits, its realization is so far in the future that we will not spend time on it here.

Machine learning: the use of computer algorithms to extract patterns from large, complex data sets with minimal human guidance

Adaptive peak: the theoretical, ideal genotype that is most fit in a given environment

Confounded: a statistical relationship where two variables are correlated so that their effects are difficult or impossible to separate

OVERVIEW

In this review, we focus on several major questions facing quantitative genetics as it transitions from Breeding 3.0 to Breeding 4.0. Crop quantitative genetics has always straddled the boundary between basic and applied science, so it is no surprise that these questions touch on basic mechanisms of evolution, domestication, and development even as they impact plant improvement practices and agronomic performance.

We have chosen three key questions to consider: (*a*) How do we adapt crops to better fit agricultural environments? (*b*) What is the nature of the diversity upon which breeding can act? (*c*) How do we deal with deleterious variants? We still have only partial answers to each of these questions, and fleshing them out is a prerequisite to fully harnessing the tools of Breeding 4.0. Although many of our examples come from the grasses owing to the authors' familiarity with them, these issues apply across all crops.

HOW DO WE ADAPT CROPS TO BETTER FIT AGRICULTURAL ENVIRONMENTS?

At the risk of stating the obvious, crops did not evolve in agricultural settings. Modern agriculture puts plants in a very different environment from their ancestors, with many selective pressures operating in different ways (**Figure 2**). As a result, new alleles are pushed to high frequencies, and the resulting selective sweeps have been found in maize (72), wheat (18), rice (59, 190), soybean (100, 202), tomato (84, 102), and sunflower (20), among others. Importantly, no crop has yet reached the top of its adaptive peak, especially since changing agronomic practices keep moving it. Finding ways to move crops closer to their peak is the ultimate goal of modern plant breeding. We focus here on three aspects of agriculture where increased understanding is likely to benefit breeding in the near term: new mega-environments, interplant competition, and soil interactions.

New Mega-Environments

A mega-environment is a group of environments sharing similar climatic conditions (day length, rainfall, temperature, etc.). Most crops originated in very localized regions but were spread by humans throughout the world (82). Adaptations to new mega-environments are thus among the largest changes crops have gone through since domestication.

The fundamental problem with finding the genes involved in these adaptations, however, is that population structure is usually highly confounded with the same adaptations one wants to map (160). Although statistical methods can reduce false positives due to background population structure (137), the correlation between adaptation and structure usually means they also reduce true positives. Such confounding can be mitigated by decoupling population structure and adaptation, either through experimental design or by selecting populations where they are naturally decoupled.

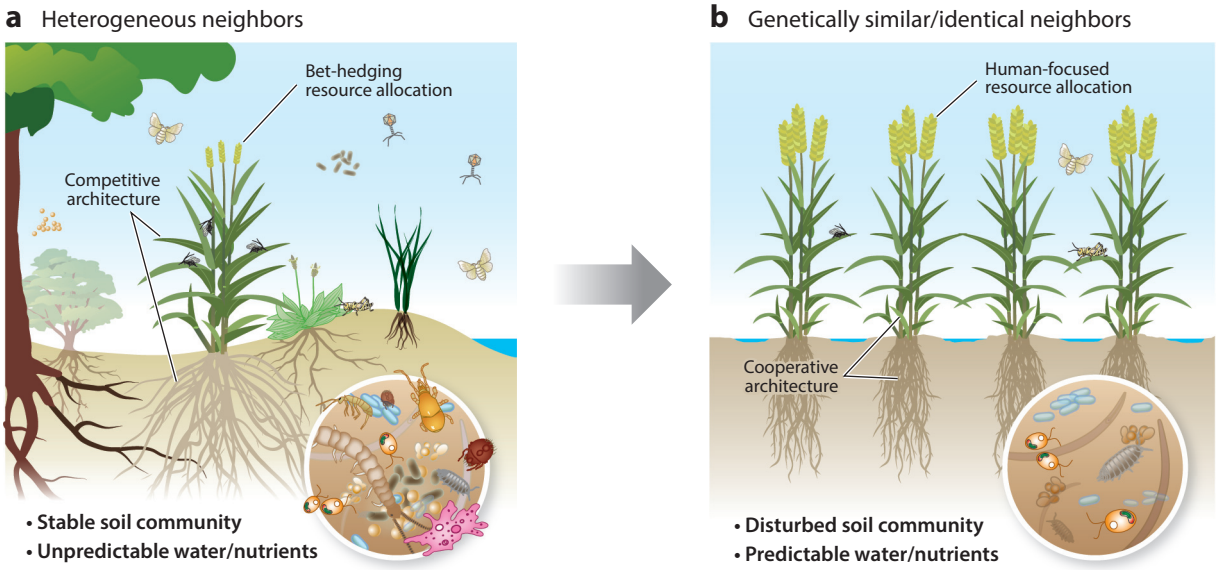

Figure 2

Plant adaptive context. (*a*) The environment in which crop wild ancestors evolved is very different from (*b*) the environment in modern agriculture. Some of the changes involved are relatively well studied, such as repartitioning resources from vegetative tissue to grain and fruit. Others, such as how plants evolve to live in a perpetually disturbed soil ecosystem, are only beginning to be understood.

In agriculture, the gold standard experimental designs to study adaptation usually follow either a Multiparental Advanced Generation Intercross (MAGIC) (87) or Nested Association Mapping (NAM) (117) design. Although the details differ, both methods separate adaptation from population structure by using a small number of parents to create a large number of recombinant progeny. The resulting population has a balanced structure and retains the statistical power of other controlled mating designs, allowing adaptive alleles to be identified with high confidence. Such analyses have been used to identify genes controlling flowering time and photoperiod sensitivity (14, 73, 107, 114, 155), drought tolerance (96), salinity tolerance (150), cold tolerance (152), and development rate (124).

An alternate approach is to take advantage of populations where structure is naturally decoupled from selection. This situation usually occurs in organisms with large population sizes that live across environmental gradients with few barriers to gene flow. Environmental genome-wide association studies using such populations have proven extremely powerful and include plant height and inflorescence architecture in sorghum (94, 122); photoperiod-, altitude-, and drought-related adaptation in maize (4); climatic adaptation in conifers (195); a host of environmental conditions in *Arabidopsis* (55, 93); and even salinity tolerance in Atlantic herring (112). Although not every organism is suitable for this sort of analysis, taking advantage of those that are can identify genes and alleles that are adaptive for different environmental conditions. Breeders can then target these genes for improvement in related organisms.

Photoperiod sensitivity: the tendency of many plants to initiate flowering based on how much light they receive each day

Interplant Competition

The wild ancestors of crops rarely, if ever, existed in genetically homogenous stands, and even traditional landraces are usually heterogeneous mixes of genotypes. Modern crops, however, often

grow in genetically uniform stands where the combined performance of a field is more important than any individual plant. This imposes a form of group selection on modern cultivars, so that plant architecture is optimized for whole-field production through steeper leaf and root angles, shorter plants, better tolerance of high planting densities, and similar traits (35, 37).

Group selection: a state where selection acts on the fitness of the entire group instead of individuals

A competitive plant is one that is good at taking resources for itself at the expense of its neighbors. Although several authors have noted that less-competitive plants yield better in modern agricultural settings (e.g., 35, 127, 151), the genetic loci controlling competitiveness have only rarely been mapped. This may be because the basic traits that underlie competition (root angle, leaf angle, metabolite production) are usually interesting in their own right, so they are often mapped without regard for competition (11, 16, 197). The few studies directly mapping competition confirm that the beneficial loci in low-competition settings are different from ones in high-competition settings (90, 166, 196). Especially at early breeding stages, both breeding and research environments tend to reward plants that strongly compete with their genetically different neighbors, even though these plants do not consistently deliver the best yields when planted in uniform stands (41). Integrating selection for lower competition into most breeding pipelines requires either alternative field schemes or (more likely) in silico modeling (203) to determine optimal trait combinations to include in breeders' selection indices.

Soil Interactions

Soil is not just a substrate for crop growth; it is arguably one of the most complex ecosystems on Earth (30). Repeated tilling leaves the soil in a state of perpetual disturbance (both physically and biologically) with highly altered microbial profiles that are further shifted by fertilizers, irrigation, crop rotation, and other agronomic practices (33, 57, 58, 113, 128, 162, 184). Such disturbance creates a situation where plants' evolved responses are maladapted to the ecology of modern agricultural soil. Nascent research seeks to understand the genetics of microbial associations, such as in recruiting beneficial rhizobia (2, 29, 80, 164) and mycorrhizae (2, 95), establishing microbial communities (3, 19, 174), excluding pathogens (5, 118), and potentially influencing food quality (12). Microbial communities can alter host phenotypes (132, 179), although very little is known about the quantitative genetics involved. Microbial profiling and metagenomic analyses of environmental (170) and crop-associated (40, 49, 135) microbial communities could open a window onto how to harness these communities for future breeding efforts. It should be noted, however, that research in several plants indicates that the environment is probably the biggest driver of plant microbial communities (77, 106, 135), especially in the rhizosphere. This means it may be easier to adjust management practices or apply specific microbes directly to seeds or plants than to breed for improved associations.

WHAT IS THE NATURE OF THE DIVERSITY UPON WHICH BREEDING CAN ACT?

Although breeding can treat genetics like a black box (and did so in Breeding 1.0 and much of 2.0), understanding the nature of plant genetic variation can significantly improve breeding efforts. One of the best tools for this is the plant's own genome because it contains a history of environmental adaptation. Comparing different genomes within a species thus gives clues to which loci are important; this sort of analysis is especially powerful in crops with hundreds to thousands of genomes available (e.g., 1, 15).

A major conclusion of whole-genome comparisons is that plant variation is significantly driven by variation in gene content and copy number instead of just differences in protein sequences.

This hypothesis was first put forward nearly 20 years ago (154) and suggested that the presence and absence of genes might underlie important phenomena. such as hybrid vigor (46). Presence–absence variation explains significant amounts of phenotypic variation (22, 104, 180), although many presence–absence genes have low RNA expression (68) and even fewer of them are translated into proteins (182). It thus seems that, even though many genes can have variations in copy number, not all such variations are important. Most of the important presence–absence variation in plants appears to stem from either polyploidy or tandem duplications, as opposed to other mechanisms like transposon duplication.

Polyploidy

There have been at least two major polyploid events in all angiosperms, and many more in most lineages (79, 186). Although this implies that polyploidy is the norm, most polyploid events are actually evolutionary dead ends (115), and it is only the lucky few that survive in the long term.

Even though plants readily undergo polyploidization, the fitness consequences of polyploidy are still not completely understood (108). One known consequence is that many of the duplicated genes are rapidly lost via mutation, deletion, or other mechanisms. This loss is not random; rather, one syntenic segment (subgenome) from a given region tends to remain relatively intact, whereas the other is preferentially mutated, deleted, or otherwise degraded (156) (**Figure 3**). Both maize and *Brassica rapa*, for example, show clear evidence of differential genome retention (21, 145). Bread wheat (*Triticum aestivum*) is an example of this process in action, as a substantial portion of its allelic variation stems from dosage-altering loss-of-function mutations stemming from its polyploidy approximately 10,000 years ago (88). Subgenome dominance is apparently established very rapidly with both gene expression differences and epigenetic marks of dominance appearing in the very first generation (39).

Tandem Duplication

Tandem duplication of genes occurs because of replication errors or unequal crossing-over during meiosis (143). Since tandemly duplicated genes are often still under the control of their native regulatory elements, these duplications provide an easy mechanism for modification of gene dosage. For example, variation in carotenoid degradation in maize is due to a dioxygenase gene that first transposed approximately 2 Mb away and then was tandemly duplicated up to 23 times, providing significant quantitative variation in carotenoid degradation (168). Tandem duplication has also been implicated in aluminum tolerance in maize (109), salt tolerance in wheat (200), and dwarfing in both wheat (99) and sorghum (125).

The Size of Mutational Space

Given that both polyploidy and tandem duplication can lead to large increases in gene content (including significant variation within a species), how big is the actual mutational space in a crop? Or in other words, how many different mutations can shift a crop toward the same goal? Although population genetics highlight the opportunities for convergent evolution (140), older studies suggest the space for mutations can be remarkably small. For example, mutations in numerous genes can produce the sweet corn phenotype, but three of four independent evolutions were all in one cleft of a single enzyme (172). Meanwhile, rice aromas have developed via 10 independent mutations in the same gene (86), and the nonshattering phenotypes of domestic sorghum, rice, and maize are all due to independent mutations in the same orthologous genes (103). These were all recent

Presence–absence variation: genetic variation in the form of genes that are present in one individual but completely missing in another

Polyploidy: the state of an organism having more than two complete genome copies

Tandem duplication: when one or more copies of a gene are inserted next to the original

Transposon duplication: the copying of a gene by its hitchhiking with a transposon that is copying itself to a new location

Syntenic: corresponding regions in two genomes that have similar gene structure due to sharing a common ancestor

Gene dosage: the number of copies of a gene in an organism, especially relative to other genes in the same pathway

Convergent evolution: the ability of two organisms to independently evolve similar or identical traits

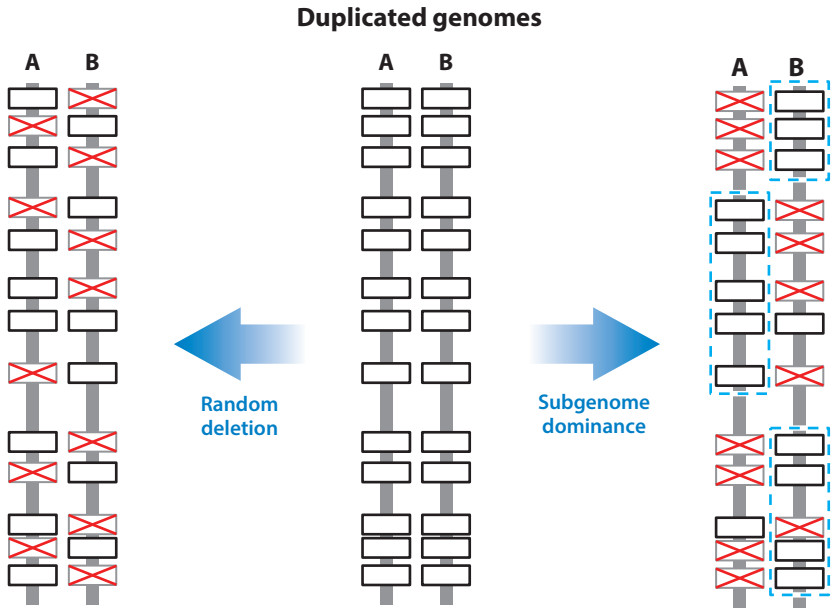

Figure 3

Subgenome dominance. Immediately after a polyploidy event, the genome contains two complete sets of genes (*A* and *B*, *center*). Over time, many genes are reduced to a single copy owing to mutations and deletions (*red Xs*). Although one would expect this process to be random (*left*), most genomes show evidence of genes that were preferentially retained or lost in blocks (*right*). The dominant subgenome consists of the blocks that are preferentially retained (*blue boxes*). Note that which subgenome a block belongs to depends only on how well it is retained, not on where it came from. In other words, a species with both A and B genomes due to polyploidy can have a dominant subgenome consisting of a mix of A and B segments, instead of A being completely dominant to B or vice versa.

selections caused by humans, but natural variation follows the same pattern. For example, sorghum and maize diverged about 12 million years ago, during which time maize underwent a whole-genome duplication from which it currently retains 20–30% of the duplicated genes (156). Despite this distance, large mapping projects in both species show consistent alignment of the quantitative trait loci (QTLs), where one locus in sorghum is matched by two syntenic QTLs in maize (107). All of this indicates that, in practice, there are only a limited number of routes to alter a given phenotype.

What about raw genome size? Angiosperm genomes can vary over three orders of magnitude (8), from the carnivorous herb *Genlisea tuberosa* (0.061 GB) (44) to the canopy plant *Paris japonica* (149 GB) (136). At first glance, these differences in genome size can appear to have strong impacts on QTL, especially ones that act through gene regulation. For example, in *Arabidopsis* (0.125–0.150 GB genome), almost all QTLs are within 5 kb of the affected gene (185). By contrast, two of the best-characterized QTLs in maize (~2.6 GB genome)—*teosinte branched 1* and *vegetative to generative transition 1*—are in enhancer elements approximately 60 kb away from the genes they affect (24, 153). However, even though the distance between important DNA elements can vary dramatically across plant genomes, the actual genomic space that is important seems to be much smaller and more constant. For example, nearly 90% of maize phenotypic variation could be assigned to the 3% of the genome that is either protein-coding or noncoding open chromatin (149). This gives a much smaller sequence space to search for variation. Similarly, only 5–7%

Quantitative trait loci (QTLs): locations in the genome that have been identified as influencing a quantitative trait

of the rice genome is in open chromatin regions that are likely to impact function (199). If this pattern holds across other plants, it implies that the functional mutational space may be quite small indeed.

Haplotype: a set of alleles linked together on the same piece of DNA

HOW DO WE DEAL WITH DELETERIOUS VARIANTS?

Human geneticists often focus on deleterious mutations because of their role in diseases, but most plant breeders do not think about them much. Any obvious deleterious mutation is eliminated early, with little thought given to molecular mechanisms or how many milder mutations go unseen. Although *Arabidopsis* is estimated to accumulate 1 mutation per generation (131), maize appears to accumulate nearly 90 of them (23). Assuming 5–10% of the genome has a functional role (149), this means five to nine mutations per generation could potentially affect a phenotype, and most of them would probably be detrimental. Identifying, controlling, and repairing these mutations is likely to be a major aspect of research as we move toward Breeding 4.0.

Deleterious Alleles and Domestication

When researchers compare domesticated species with their wild relatives, they find that domestication is often associated with an increase in the number of apparently harmful alleles. This pattern can be seen in both plants and animals, including rice (105, 126), sunflower (144), tomato (84), dogs (110), and horses (159). On the basis of population genetic theory, there are a few processes that may explain this increase in detrimental alleles. First, domestication alters selection pressures so that traits favored in the wild become neutral or disfavored under domestication (74). Second, purifying selection is frequently reduced after domestication bottlenecks (83). Third, inbreeding of domesticated varieties further reduces their effective population size and effective recombination rates (105). Although the relative contribution of each force varies by species, all of them probably act to some extent on domesticated crops.

Emerging evidence suggests that this cost of domestication can be tempered by modern improvement practices. For example, modern inbred lines of maize carry fewer nonsense mutations than their wild relatives (22) and have less genetic load than traditional outcrossing landraces (193). Young alleles generally appear to be under more stringent purifying selection in maize than in its ancestor, teosinte (7). In cassava, meanwhile, domestication loci have fewer deleterious mutations in cultivated varieties compared to progenitor populations even as drift has increased plants' overall genetic load (141).

The Problem with Chromosomes

One problem with deleterious variants is that the biology of chromosomes makes them hard to breed out. Organizing genes into large linear chromosomes helps cells properly segregate complete genomes during division, but this structural constraint also reduces the efficiency of selection by linking alleles that may have different (and often opposite) fitness values. This means that any given individual has a mix of good and bad alleles linked together within a haplotype, and recombination is often not efficient enough to ever create a single ideal combination. Instead, multiple suboptimal haplotypes selectively interfere with one another so that none reach fixation. This phenomenon is called Hill-Robertson interference (65) and results in negative linkage disequilibrium (repulsion phase) between beneficial variants in low-recombination regions (42).

Hill-Robertson interference has important (but often overlooked) consequences for plant breeding. First, interference reduces the efficacy of selection on any individual site (65). This

allows slightly deleterious variants to accumulate and reduces the probability of fixation for any given beneficial allele, a process that has been shown in rice (105), maize (148), and sunflower (144). Second, having beneficial alleles spread across multiple nonrecombining haplotypes reduces genetic variance by eliminating extreme phenotypes, which in turn decreases the raw material for selection to act upon. Third, assuming most deleterious alleles are at least partly recessive, the low-recombination regions where interference is most severe should benefit more from heterozygosity than high-recombination regions. This benefit arises because low-recombination regions have a greater chance to complement deleterious alleles. Artificial inbreeding should thus favor individuals that retain heterozygosity in low-recombination regions, as has been verified in oat (71) and maize (117). Interference and its consequences are easiest to see around centromeres because they are so large, but the same processes also occur in localized regions across the genome.

Heterosis: the tendency of a hybrid offspring to be more fit than either of its inbred parents

One can assume that at least some QTLs of agronomic importance are located in similar low-recombination regions, making their improvement extremely difficult. One work-around is to select for haplotypes that complement each other in the hybrid state, which is what results in distinct heterotic groups in maize and other hybrid breeding programs (169). Looking toward Breeding 4.0, technologies that allow breeders to precisely manipulate recombination sites or alter specific alleles through genome editing could bypass the Hill-Robertson effect entirely.

Implications for Heterosis

Heterosis has been known for 150 years (31), but its molecular underpinnings are still debated (157). Because of the complementation mentioned earlier, regions with strong Hill-Robertson interference are expected to behave as a single overdominant locus (98) even though the individual alleles are strictly dominant. This effect is called pseudo-overdominance.

Experimental work confirms that at least some apparently overdominant loci are actually pseudo-overdominant. Several generations of random mating in a biparental maize family eliminate the appearance of overdominance (48, 121), and grain QTLs consistently localize to the centromere and pericentromere (92, 158), which are both predicted to show strong pseudo-overdominance. Confirming this prediction requires fine-mapping traits, something that by definition is difficult to do in low-recombination regions but has been managed occasionally. For example, a single overdominant QTL near the centromere of maize chromosome 5 could be separated into two dominant QTLs in repulsion phase (53). Meanwhile, a sorghum height QTL separated into two genes with opposite effects approximately 3 Mb apart (98); both alleles were dominant and resulted in apparent overdominance in the hybrid.

Although we have focused on low-recombination regions, these regions usually have relatively few genes in them (6, 38, 75). Therefore, even if the most obvious genetic load is in low-recombination regions, the most overall load is likely in high-recombination regions. Selection is more efficient in these regions, but other factors associated with high recombination can oppose movement up the fitness landscape. Most notably, GC-biased gene conversion likely maintains some deleterious mutations at frequencies much higher than expected under mutation-selection-drift equilibrium (52, 149).

The importance of low-recombination regions for heterosis may also depend on the species. In bread wheat, for example, heterosis appears to be more driven by epistatic effects than by dominance effects (78). Significant heterosis has also been seen from gene-dosage effects in rice (70), maize (194), and *Arabidopsis* (45), indicating that the relative amounts of different genes are just as important as dominant and recessive relationships. Heterosis also changes in different environments (101), which makes it more difficult to pin down consistent mechanisms.

DELIVERING QUANTITATIVE GENETICS TO THE FIELD

How important is it to understand the adaptive and deleterious variants across the genome? This sort of knowledge has already proven useful in a few specific pathways, especially for traits that are difficult and/or expensive to score and that are controlled by a small number of genes. Such traits include many disease resistance loci, submergence tolerance in rice (189), and carotenoid accumulation in maize (56, 192). For Breeding 3.0 and 4.0 to achieve their full potential, we must identify how to translate basic knowledge into real-world results.

Fisher-Orr and the Diminishing Returns of Breeding

The basic unit of modern crop breeding is the QTL. The more beneficial QTL you can breed into a variety, the better that variety becomes and the higher up the adaptive peak it climbs. First-generation QTL mapping focused on the big domestication loci (e.g., 13, 85, 103, 139, 165, 167, 183) and disease resistance genes (e.g., 69, 111, 163, 191). These efforts were very successful, but as time passed, the pattern emerged that newer QTLs generally had smaller effect sizes, especially in outcrossing species. Smaller QTLs require more effort to map and provide less benefit, putting crop breeding in a state of diminishing returns.

Several phenomena contribute to this pattern. First, scientists were rational: They cloned biggest QTL first. But another major factor was that most QTL effects appear to follow the Fisher-Orr geometric model (129, 130). In brief, this model says (*a*) large-effect mutations are more likely to be harmful than helpful and (*b*) selection is less efficient for complex traits than for simple ones. In breeding contexts, this means that alleles with large beneficial effects rapidly become fixed, after which only successively smaller and smaller allele effects actually move plants closer to the adaptive peak (**Figure 4**). In addition, most beneficial alleles for complex traits (e.g., yield) have small effects in the first place.

Large mapping panels have confirmed that most alleles have small effects (14, 116, 133, 180). This implies that, especially in major crops, many of the large-effect QTLs have probably already been identified and fixed in elite germplasm, leaving breeders to work with progressively smaller effects that are harder to identify. This appears to be especially true in outcrossing species like maize (14, 134) and even cattle (reviewed in 64). Self-pollinating species appear to have more large-effect genes, such as the dwarfing genes of the Green Revolution (60), flowering time in barley (114), and drought tolerance in rice (178). It may be that in the constant reshuffling of haplotypes, each generation in outcrossing species selects for small-effect alleles that play nice with other genes. By contrast, self-pollinating crops pass down complete haplotypes to their offspring, which may more readily allow large-effect alleles to evolve.

Genomic Prediction

One method of dealing with increasingly small QTLs is to stop trying to map them and instead work with the entire genome. This is the approach of genome-wide prediction and genomic selection (**Figure 5**), a pair of incredibly successful paradigms that are affecting almost all sectors of breeding (63). Genome-wide prediction is the process of using genetic data to predict an individual's phenotype; genomic selection is simply using those predictions to make breeding decisions. This sort of selection scheme was first demonstrated by Meuwissen et al. (119), and its goal is to improve breeding by both reducing the cost of phenotyping and reducing the cycle time for early generation selections.

Genomic selection has proven extremely beneficial to dairy cattle, and signs indicate it will be similarly valuable for crop breeding (28, 63, 176). However, there are still many unknowns, such

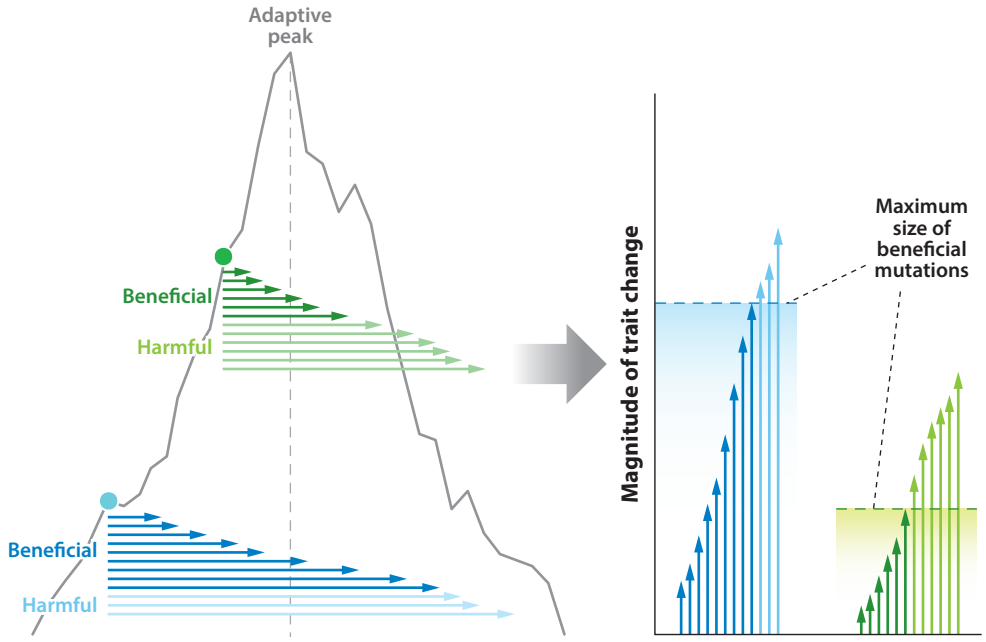

Figure 4

The Fisher-Orr model. The Fisher-Orr model states that the closer an organism is to an adaptive peak, the more likely a large mutation is to be harmful. The blue individual is far from the peak, so many different mutations result in a net benefit. By contrast, the green individual is close enough to the peak that even moderately sized changes can overshoot the ideal and leave its fitness lower than before.

as the optimal allocation of resources or choice of experimental designs. Krchov & Bernardo (89), for example, found that some breeding designs were always better under genomic selection, some were always better under phenotypic selection, and many could shift between the two based on budget, resource allocation, trait heritability, and other factors. Heffner et al. (61) estimate that genomic selection outperforms marker-assisted selection in terms of gain per unit time as long as the trait has a heritability of 0.2 or greater. Meanwhile, the choices of statistical models for genomic selection are many and growing (51); although with real-world data, there seem to be little practical differences among them (62). Genomic selection for hybrid performance is also becoming more powerful (76, 97), which is likely to make genomic selection of hybrids possible even for historically inbred crops like wheat (201).

Genomic selection has great potential for improving crops, but in most cases, that potential is still unrealized. This is most true for species outside of the major row crops and outside of the developed world, largely due to simple logistics. Genomic prediction requires high-throughput genotyping systems and bioinformatics expertise that many programs do not have, and breeding decisions for annual crops must occur in two to three brief windows for genomic selection to be worthwhile. The cost of genotyping can also be a major determinant of whether genomic selection is worthwhile (89, 146), though the still-falling price of sequencing implies that may not be true for long. Prediction models are also limited to highly related germplasm; models using different breeding programs rarely have value, and accuracies can rapidly drop even beyond half-sib family structures (10). All of these barriers need to be overcome for genomic selection to see truly global deployment.

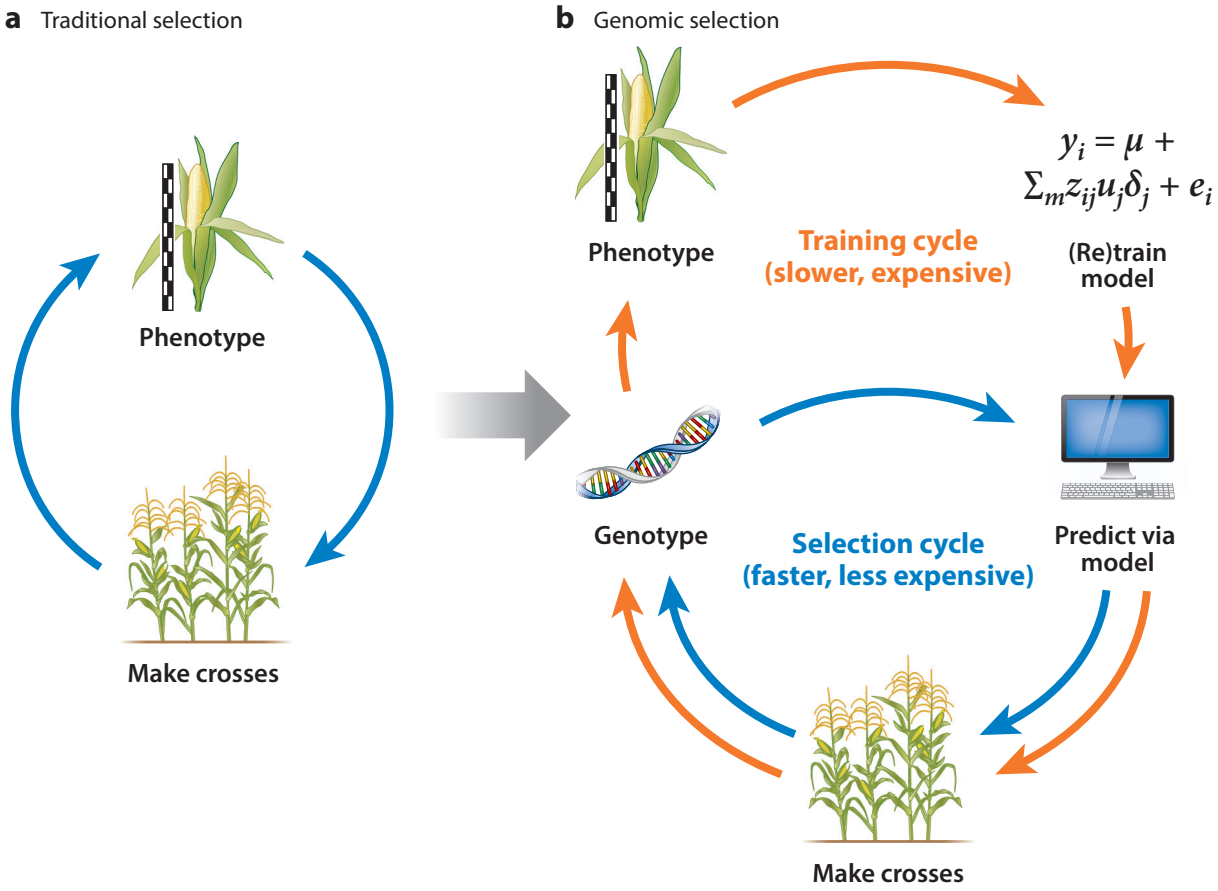

Figure 5

Genomic selection. (*a*) Traditional selection involves cycles of making crosses to get new genetic combinations, evaluating the new varieties by phenotyping, and using those evaluations to select the next generation of parents. (The point at which a variety is spun off into production is not shown.) (*b*) Genomic selection adds several layers to this scheme and splits the process into training and selection cycles. When training, the breeder must go around the entire outside loop: making crosses, getting genotype and phenotype information from them, building a mathematical model, and finally using that model to select future parents. After that, the breeder can skip the training portion and go directly from new material to genotype to choose parents based on the model, with no need to phenotype. Although the model must be retrained every few generations, breeders generally save significant time and money by skipping the (usually laborious and expensive) phenotyping step two or three times.

Democratizing Breeding 3.0

Many crop breeding programs are still at the Breeding 1.0 or 2.0 stages, especially in the developing world, yet widely deploying Breeding 3.0 is critical to providing benefits for the global population. This democratization process is already underway, as inexpensive genotyping catapults formerly neglected crops into twenty-first century genetics. Millets (9, 175, 177, 181), cassava (138), cacao (123), strawberry (67, 161), sweet potato (66), and many others are benefiting from this technology, but it must reach even further to have truly global benefits. DNA sequencing is likely to drop to $1/Gb in the next several years, which implies that virtually any breeding line can be skim sequenced for little cost. High-throughput sequencing facilities, meanwhile, will probably reduce the price of genotyping seeds to only $1 or $2 per sample. However, while sequencing and

genotyping are dropping to negligible prices, field work and the logistics of tracking and processing thousands of samples are not. New ideas and informatics systems are needed to enable small programs to deal with these logistics. Most breeders do not have the time or skills to process the data themselves, and ideally breeders should never even need to look at a nucleotide sequence. Projects like the Genomic Open-source Breeding Informatics Initiative are starting this effort, but a larger international community of both public and private entities is needed to make Breeding 3.0 a global reality.

MOVING TO BREEDING 4.0

Beyond just democratizing Breeding 3.0, what do we need to move to Breeding 4.0 and capitalize on the ability to manipulate individual genes and even single base pairs? The last decade has seen tremendous strides for identifying functional variants through genome-wide associations, evolutionary constraint, comparative genomics, and a range of molecular biology assays. For breeding, however, each of these approaches needs to be carefully measured in terms of its usefulness. Unlike human genetics, crops do not have one focus species but rather hundreds of species, each with many different breeding targets around the globe. A major question for the next decade is how to leverage knowledge across these species and breeding programs to make the most use of limited resources. For example, how can maize breeding learn from rice trials, what lessons can cassava take from apple or cotton breeding, and how much of what is learned in *Arabidopsis* can we apply to bananas, blueberries, oats, or loblolly pine?

A detailed understanding of every gene's function is probably not needed for Breeding 4.0. We do, however, need to be able to estimate of the effect of each functional variant in a range of target environments. On a practical level, how do we get there? Although a species could have 50 million common variable sites in the genome, most of them probably do not matter very much. To reach Breeding 4.0, we need to reduce this sea of variation to a few tens of thousands of high-probability functional sites.

What types of data are the most useful for this filtering? We present some recommendations below:

1. Genetic mapping of complex target traits is by far the most expensive but also the most important approach to identifying functional variants. Complex traits such as yield rarely resolve to single genes, and trying to resolve any given QTL to a single nucleotide is rarely worth the cost. Instead, the central goal of mapping should be to resolve a few key QTLs to the gene level to highlight pathways and processes not previously considered. However, the greatest value of genome-wide association studies and genetic mapping may well be the data sets of genotypic and phenotypic data they produce. These can be used to benchmark every other approach, especially in the context of multiple environments, and can be integrated into ongoing breeding efforts.
2. Genomic annotation is the process of distilling our molecular understanding of the genome down to machine processable data. It includes annotations with protein domains, gene ontologies, methylation patterns, chromatin states, transcription factor binding sites, expression levels, and many others. Although all of these measures may be biologically interesting, there are very strong correlations among them, and many can be predicted from each other. The community needs to rigorously define the cost-effectiveness of each annotation to identify where to best put limited resources.
3. The mapping of intermediate phenotypes—for example, RNA transcripts or metabolite abundance—gives much higher resolution than the mapping of terminal phenotypes such

as yield and frequently identifies the exact gene involved or even specific causal variants. Related approaches such as chromatin profiling and proteomics could dramatically improve our ability to identify functional variants. The key challenge is to make these technologies cheap enough that crops can be profiled across a range of genotypes and environments.

4. Evolution is the ultimate yield trial, as it integrates the success of various genotypes over millions of years. The greatest limitation of current approaches is that sequencing distant evolutionary species only provides information on the most conserved elements. Saturation of closely related species is needed to fully understand regulatory conservation. Although crops are grown in environments where they did not evolve, the careful choice of landraces or wild analogs (as opposed to elite lines) may still be able to address these questions.

Linkage drag: the process whereby introgressing a desirable allele into breeding lines brings along many undesirable alleles linked to it

Each of these approaches can enrich for functional variants but not identify them for certain. Intelligent integration of these approaches, however, could provide whole sets of high-confidence variants in a cost-effective manner. This integration requires characterizing global germplasm resources both genetically and phenotypically, developing informatics tools to share this information, and using appropriate methods (e.g., machine learning or similar) to integrate these diverse data sets.

A final key to Breeding 4.0 is large-scale genomic editing (tens to hundreds of sites per generation). Direct genome editing will then almost certainly replace crosses as the most efficient way to tailor genetic variation into optimal combinations. Better still, it can do so without linkage drag destabilizing the decades of breeding that went before. Of course, such edited crops would probably need to overcome consumer resistance to engineered foods, but that is an entirely different aspect to global food security.

Breeding has always been a numbers game, so we do not need to identify every variant with 100% accuracy. Even if we are only right 10% of the time, it may be enough to push crop breeding faster and more cost-effectively than we could otherwise. The hunt for these variants is already underway in many labs around the world. Integrating all this work into breeding pipelines will be key to providing an adequate, nutritious, and sustainable food supply for the entire globe throughout the twenty-first century.

SUMMARY POINTS

1. Breeding can be divided into four stages based on available technology; we are currently in Breeding 3.0.
2. To reach Breeding 4.0, we need to identify specific alleles that are responsible for desirable variation in crops.
3. Crops are still only partly adapted to agriculture and can be improved by breeding them to better fit modern growing environments.
4. Variation that is relevant to agriculture is not randomly spread through the genome, so finding the relevant portions can improve the focus of breeding efforts.
5. Much of Breeding 4.0 probably revolves around identifying and purging deleterious variants.
6. More effort is needed to democratize current and future breeding technologies so that they benefit all of global agriculture.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We wish to thank Sara Miller for help preparing the manuscript, and the United States Department of Agriculture-Agricultural Research Service, the University of Georgia, and DowDuPont for funding support that made the preparation of this article possible.

LITERATURE CITED

1. 3,000 Rice Genomes Project. 2014. The 3,000 rice genomes project. *Gigascience* 3:7
2. Afkhami ME, Stinchcombe JR. 2016. Multiple mutualist effects on genomewide expression in the tripartite association between *Medicago truncatula*, nitrogen-fixing bacteria and mycorrhizal fungi. *Mol. Ecol.* 25(19):4946–62
3. Agler MT, Ruhe J, Kroll S, Morhenn C, Kim S-T, et al. 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLOS Biol.* 14(1):e1002352
4. Romero Navarro JA, Wilcox M, Burgueño J, Romay C, Swarts K, et al. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* 49:476–80
5. Álvarez-Pérez JM, González-García S, Cobos R, Olego MÁ, Ibañez A, et al. 2017. Use of endophytic and rhizosphere actinobacteria from grapevine plants to reduce nursery fungal graft infections that lead to young grapevine decline. *Appl. Environ. Microbiol.* 83(24):e01564-17
6. Anderson LK, Lai A, Stack SM, Rizzon C, Gaut BS. 2006. Uneven distribution of expressed sequence tag loci on maize pachytene chromosomes. *Genome Res.* 16(1):115–22
7. Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J. 2016. Recent demography drives changes in linked selection across the maize genome. *Nat. Plants* 2:16084
8. Bennett MD, Leitch IJ. 2012. Angiosperm DNA C-values database (release 8.0, Dec. 2012). *Kew R. Bot. Gard.* **http://data.kew.org/cvalues/**
9. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, et al. 2012. Reference genome sequence of the model plant Setaria. *Nat. Biotechnol.* 30(6):555–61
10. Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, et al. 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop. Sci.* 55:154–63
11. Biscarini F, Cozzi P, Casella L, Riccardi P, Vattari A, et al. 2016. Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. *PLOS ONE* 11(5):e0155425
12. Bokulich NA, Thorngate JH, Richardson PM, Mills DA. 2014. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *PNAS* 111(1):E139–48
13. Brewer MT, Moyseenko JB, Monforte AJ, van der Knaap E. 2007. Morphological variation in tomato: a comprehensive study of quantitative trait loci controlling fruit shape and development. *J. Exp. Bot.* 58(6):1339–49
14. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, et al. 2009. The genetic architecture of maize flowering time. *Science* 325(5941):714–18
15. Bukowski R, Guo X, Lu Y, Zou C, He B, et al. 2015. Construction of the third generation *Zea mays* haplotype map. *GigaScience* 7(4):gix134
16. Burridge J, Jochua CN, Bucksch A, Lynch JP. 2016. Legume shovelomics: high—throughput phenotyping of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* subsp, *unguiculata*) root architecture in the field. *Field Crops. Res.* 192:21–32
17. Cai W, Borlace S, Lengaigne M, van Rensch P, Collins M, et al. 2014. Increasing frequency of extreme El Niño events due to greenhouse warming. *Nat. Clim. Chang.* 4(2):111–16

18. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, et al. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *PNAS* 110(20):8057–62
19. Chaparro JM, Badri DV, Vivanco JM. 2014. Rhizosphere microbiome assemblage is affected by plant development. *ISME J.* 8(4):790–803
20. Chapman MA, Pashley CH, Wenzler J, Hvala J, Tang S, et al. 2008. A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell* 20(11):2931–45
21. Cheng F, Wu J, Liu B, Wang X. 2015. Genome evolution after whole genome triplication: the subgenome dominance in *Brassica rapa*. In *The Brassica rapa Genome*, ed. X Wang, D Kole, pp. 107–14. Berlin/Heidelberg: Springer
22. Chia J-M, Song C, Bradbury PJ, Costich D, de Leon N, et al. 2012. Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* 44(7):803–7
23. Clark RM, Tavaré S, Doebley J. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Mol. Biol. Evol.* 22(11):2304–12
24. Clark RM, Wagler TN, Quijada P, Doebley J. 2006. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* 38(5):594–97
25. Cook BI, Smerdon JE, Seager R, Coats S. 2014. Global warming and 21st century drying. *Clim. Dyn.* 43(9–10):2607–27
26. Correns CE. 1900. G. Mendel's Regel über das Verhalten der Nachkommenschaft der Rassenbastarde. *Ber. Dtsch. Bot. Ges.* 18:158–67
27. Coumou D, Robinson A. 2013. Historic and future increase in the global land area affected by monthly heat extremes. *Environ. Res. Lett.* 8(3):034018
28. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, et al. 2017. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22(11):961–75
29. Curtin SJ, Tiffin P, Guhlin J, Trujillo DI, Burghardt LT, et al. 2017. Validating genome-wide association candidates controlling quantitative variation in nodulation. *Plant Physiol.* 173(2):921–31
30. Daniel R. 2005. The metagenomics of soil. *Nat. Rev. Microbiol.* 3(6):470–78
31. Darwin C. 1868. *The Variation of Animals and Plants Under Domestication*. London: Murray
32. Darwin C. 1876. *The Effects of Cross and Self-Fertilisation in the Vegetable Kingdom*. London: Murray
33. de Quadros PD, Zhalnina K, Davis-Richardson A, Fagen JR, Drew J, et al. 2012. the effect of tillage system and crop rotation on soil microbial diversity and composition in a subtropical Acrisol. *Diversity* 4(4):375–95
34. De Vries H. 1900. Sur la loi de disjonction des hybrides. *Comptes Rendus Acad. Sci.* 130:845–47
35. Denison RF. 2015. Evolutionary tradeoffs as opportunities to improve yield potential. *Field Crops. Res.* 182:3–8
36. Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127(7):1309–21
37. Duvick DN. 2005. The contribution of breeding to yield advances in maize (*Zea mays* L.). *Adv. Agron.* 86:83–145
38. Dvorak J, Yang Z-L, You FM, Luo M-C. 2004. Deletion polymorphism in wheat chromosome regions with contrasting recombination rates. *Genetics* 168(3):1665–75
39. Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, et al. 2017. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29(9):2150–67
40. Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, et al. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. *PNAS* 112(8):E911–20
41. Fasoula VA. 2013. Prognostic breeding: a new paradigm for crop improvement. In *Plant Breeding Reviews*, ed. J Janick, pp. 297–347. Hoboken, NJ: Wiley
42. Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* 78(2):737–56
43. Fisher RA. 1919. XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ. Sci. Trans. R. Soc. Edinb.* 52(2):399–433

44. Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, et al. 2014. Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Ann. Bot.* 114(8):1651–63
45. Fort A, Tuteja R, Braud M, McKeown PC, Spillane C. 2017. Parental-genome dosage effects on the transcriptome of F1 hybrid triploid embryos of *Arabidopsis thaliana*. *Plant J.* 92(6):1044–58
46. Fu H, Dooner HK. 2002. Intraspecific violation of genetic colinearity and its implications in maize. *PNAS* 99(14):9573–78
47. Galton F. 1886. Regression towards mediocrity in hereditary stature. *J. Anthropol. Inst. Great Br. Ireland* 15:246–63
48. Gardner CO, Lonnquist JH. 1959. Linkage and the degree of dominance of genes controlling quantitative characters in maize. *Agron. J.* 51(9):524–28
49. Gdanetz K, Trail F. 2017. The wheat microbiome under four management strategies, and potential for endophytes in disease protection. *Phytobiomes* 1(3):158–68
50. Gerland P, Raftery AE, Ševcíková H, Li N, Gu D, et al. 2014. World population stabilization unlikely this century. *Science* 346(6206):234–37
51. Gianola D. 2013. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* 194(3):573–96
52. Glémin S. 2010. Surprising fitness consequences of GC-biased gene conversion: I. Mutation load and inbreeding depression. *Genetics* 185(3):939–59
53. Graham GI, Wolff DW, Stuber CW. 1997. Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. *Crop. Sci.* 37:1601–10
54. Hanasaki N, Fujimori S, Yamamoto T, Yoshikawa S, Masaki Y, et al. 2013. A global water scarcity assessment under shared socio-economic pathways—Part 2: Water availability and scarcity. *Hydrol. Earth Syst. Sci.* 17:2393–413
55. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, et al. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334(6052):83–86
56. Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, et al. 2008. Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science* 319(5861):330–33
57. Hartman K, van der Heijden MGA, Wittwer RA, Banerjee S, Walser J-C, Schlaeppi K. 2018. Cropping practices manipulate abundance patterns of root and soil microbiome members paving the way to smart farming. *Microbiome* 6(1):14
58. Hartmann M, Frey B, Mayer J, Mäder P, Widmer F. 2015. Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J.* 9(5):1177–94
59. He Z, Zhai W, Wen H, Tang T, Wang Y, et al. 2011. Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLOS Genet*. 7(6):e1002100
60. Hedden P. 2003. The genes of the Green Revolution. *Trends Genet*. 19(1):5–9
61. Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME. 2010. Plant breeding with genomic selection: gain per unit time and cost. *Crop. Sci.* 50(5):1681–90
62. Heslot N, Yang H-P, Sorrells ME, Jannink J-L. 2012. Genomic selection in plant breeding: a comparison of models. *Crop. Sci.* 52(1):146–60
63. Hickey JM, Chiurugwi T, Mackay I, Powell W, Implementing Genomic Selection in CGIAR Breed. Programs Workshop Particip. 2017. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet*. 49(9):1297–303
64. Hill WG. 2014. Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. *Genetics* 196(1):1–16
65. Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* 8(3):269–94
66. Hirakawa H, Okada Y, Tabuchi H, Shirasawa K, Watanabe A, et al. 2015. Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H.B.K.) G. Don. *DNA Res*. 22(2):171–79
67. Hirakawa H, Shirasawa K, Kosugi S, Tashiro K, Nakayama S, et al. 2014. Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Res*. 21(2):169–81
68. Hirsch CN, Hirsch CD, Brohammer AB, Bowman MJ, Soifer I, et al. 2016. Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* 28(11):2700–14

69. Huang L, Brooks SA, Li W, Fellers JP, Trick HN, Gill BS. 2003. Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. *Genetics* 164(2):655–64
70. Huang X, Yang S, Gong J, Zhao Y, Feng Q, et al. 2015. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6:6258
71. Huang Y-F, Poland JA, Wight CP, Jackson EW, Tinker NA. 2014. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. *PLOS ONE* 9(7):e102448
72. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44(7):808–11
73. Hung H-Y, Shannon LM, Tian F, Bradbury PJ, Chen C, et al. 2012. *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *PNAS* 109(28):E1913–21
74. Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *PNAS* 101(29):10667–72
75. Int. Rice Genome Seq. Proj. 2005. The map-based sequence of the rice genome. *Nature* 436(7052):793–800
76. Jacobson A, Lian L, Zhong S, Bernardo R. 2014. General combining ability model for genomewide selection in a biparental cross. *Crop. Sci.* 54:895–905
77. Jiang Y, Li S, Li R, Zhang J, Liu Y, et al. 2017. Plant cultivars imprint the rhizosphere bacterial community composition and association networks. *Soil Biol. Biochem.* 109:145–55
78. Jiang Y, Schmidt RH, Zhao Y, Reif JC. 2017. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 49(12):1741–46
79. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–100
80. Kamfwa K, Cichy KA, Kelly JD. 2015. Genome-wide association analysis of symbiotic nitrogen fixation in common bean. *Theor. Appl. Genet.* 128(10):1999–2017
81. Khoshbakht K, Hammer K. 2008. How many plant species are cultivated? *Genet. Resour. Crop Evol.* 55(7):925–28
82. Khoury CK, Achicanoy HA, Bjorkman AD, Navarro-Racines C, Guarino L, et al. 2016. Origins of food crops connect countries worldwide. *Proc. R. Soc. B.* 283(1832):20160792
83. Kirkpatrick M, Jarne P. 2000. The effects of a bottleneck on inbreeding depression and the genetic load. *Am. Nat.* 155(2):154–67
84. Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, et al. 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *PNAS* 110(28):E2655–62
85. Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, et al. 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *PNAS* 104(4):1424–29
86. Kovach MJ, Calingacion MN, Fitzgerald MA, McCouch SR. 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *PNAS* 106(34):14444–49
87. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, et al. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLOS Genet.* 5(7):e1000551
88. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, et al. 2017. Uncovering hidden variation in polyploid wheat. *PNAS* 114(6):E913–21
89. Krchov L-M, Bernardo R. 2015. Relative efficiency of genomewide selection for testcross performance of doubled haploid lines in a maize breeding program. *Crop. Sci.* 55:2091–99
90. Lake L, Li Y, Casal JJ, Sadras VO. 2016. Negative association between chickpea response to competition and crop yield: phenotypic and genetic analysis. *Field Crops. Res.* 196:409–17
91. Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121(1):185–99
92. Larièpe A, Mangin B, Jasson S, Combes V, Dumas F, et al. 2011. The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). *Genetics* 190(2):795–811
93. Lasky JR, Des Marais DL, Lowry DB, Povolotskaya I, McKay JK, et al. 2014. Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 31(9):2283–96

94. Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, et al. 2015. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* 1(6):e1400218
95. Leiser WL, Olatoye MO, Rattunde HFW, Neumann G, Weltzien E, Haussmann BIG. 2016. No need to breed for enhanced colonization by arbuscular mycorrhizal fungi to improve low-P adaptation of West African sorghums. *Plant Soil* 401(1–2):51–64
96. Li C, Sun B, Li Y, Liu C, Wu X, et al. 2016. Numerous genetic loci identified for drought tolerance in the maize nested association mapping populations. *BMC Genomics* 17(1):894
97. Lian L, Jacobson A, Zhong S, Bernardo R. 2014. Genomewide prediction accuracy within 969 maize biparental populations. *Crop. Sci.* 54:1514–22
98. Li X, Li X, Fridman E, Tesso TT, Yu J. 2015. Dissecting repulsion linkage in the dwarfing gene *Dw3* region for sorghum plant height provides insights into heterosis. *PNAS* 112(38):11823–28
99. Li Y, Xiao J, Wu J, Duan J, Liu Y, et al. 2012. A tandem segmental duplication (TSD) in green revolution gene *Rht-D1b* region underlies plant height variation. *New Phytol.* 196(1):282–91
100. Li Y-H, Zhao S-C, Ma J-X, Li D, Yan L, et al. 2013. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* 14:579
101. Li Z, Coffey L, Garfin J, Miller ND, White MR, et al. 2018. Genotype-by-environment interactions affecting heterosis in maize. *PLOS ONE* 13(1):e0191321
102. Lin T, Zhu G, Zhang J, Xu X, Yu Q, et al. 2014. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46(11):1220–26
103. Lin Z, Li X, Shannon LM, Yeh C-T, Wang ML, et al. 2012. Parallel domestication of the *Shattering1* genes in cereals. *Nat. Genet.* 44(6):720–24
104. Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, et al. 2015. High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat. Commun.* 6:6914
105. Lu J, Tang T, Tang H, Huang J, Shi S, Wu C-I. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* 22(3):126–31
106. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, et al. 2012. Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488(7409):86–90
107. Mace ES, Hunt CH, Jordan DR. 2013. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor. Appl. Genet.* 126(5):1377–95
108. Madlung A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* 110(2):99–104
109. Maron LG, Guimarães CT, Kirst M, Albert PS, Birchler JA, et al. 2013. Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. *PNAS* 110(13):5241–46
110. Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, et al. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *PNAS* 113(1):152–57
111. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, et al. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262(5138):1432–36
112. Martinez Barrio A, Lamichhaney S, Fan G, Rafati N, Pettersson M, et al. 2016. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *eLife* 5:e12081
113. Mathew RP, Feng Y, Githinji L, Ankumah R, Balkcom KS. 2012. Impact of no-tillage and conventional tillage systems on soil microbial communities. *Appl. Environ. Soil Sci.* 2012:548620
114. Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, et al. 2015. Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* 16:290
115. Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, et al. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333(6047):1257
116. McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, et al. 2016. Open access resources for genome-wide association mapping in rice. *Nat. Commun.* 7:10532
117. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, et al. 2009. Genetic properties of the maize nested association mapping population. *Science* 325(5941):737–40
118. Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, et al. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 332(6033):1097–100

119. Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–29
120. Meyer RS, DuVal AE, Jensen HR. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196(1):29–48
121. Moll RH, Lindsey MF, Robinson HF. 1964. Estimates of genetic variances and level of dominance in maize. *Genetics* 49(3):411–23
122. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, et al. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *PNAS* 110(2):453–58
123. Motamayor JC, Mockaitis K, Schmutz J, Haiminen N, Livingstone D 3rd, et al. 2013. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol.* 14(6):r53
124. Müller NA, Wijnen CL, Srinivasan A, Ryngajllo M, Ofner I, et al. 2016. Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Genet.* 48(1):89–93
125. Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS. 2003. Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. *Science* 302(5642):81–84
126. Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, et al. 2014. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Mol. Ecol.* 23(9):2210–27
127. Nasseer AM, Martin JM, Heo HY, Blake NK, Sherman JD, et al. 2016. Impact of a quantitative trait locus for tiller number on plasticity of agronomic traits in spring wheat. *Crop. Sci.* 56:595–602
128. Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. 2017. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. *Appl. Environ. Microbiol.* 84:e01646-17
129. Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52(4):935–49
130. Orr HA. 2000. Adaptation and the cost of complexity. *Evolution* 54(1):13–20
131. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, et al. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327(5961):92–94
132. Panke-Buisse K, Poole AC, Goodrich JK, Ley RE, Kao-Kniffin J. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J.* 9:980–89
133. Pascual L, Desplat N, Huang BE, Desgroux A, Bruguier L, et al. 2015. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J.* 13(4):565–77
134. Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, et al. 2014. The genetic architecture of maize height. *Genetics* 196(4):1337–56
135. Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, et al. 2013. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *PNAS* 110(16):6548–53
136. Pellicer J, Fay MF, Leitch IJ. 2010. The largest eukaryotic genome of them all? *Bot. J. Linn. Soc.* 164(1):10–15
137. Price AL, Zaitlen NA, Reich D, Patterson N. 2010. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11(7):459–63
138. Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, et al. 2012. The cassava genome: current progress, future directions. *Trop. Plant Biol.* 5(1):88–94
139. Purugganan MD, Boyles AL, Suddith JI. 2000. Variation and selection at the *CAULIFLOWER* floral homeotic gene accompanying the evolution of domesticated *Brassica oleracea*. *Genetics* 155(2):855–62
140. Ralph PL, Coop G. 2015. Convergent evolution during local adaptation to patchy landscapes. *PLOS Genet.* 11(11):e1005630
141. Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, et al. 2016. Cassava HapMap: Masking deleterious mutations in a clonal crop species. BioRxiv 077123. **https://doi.org/10.1101/077123**
142. Ray DK, Mueller ND, West PC, Foley JA. 2013. Yield trends are insufficient to double global crop production by 2050. *PLOS ONE* 8(6):e66428
143. Reams AB, Roth JR. 2015. Mechanisms of gene duplication and amplification. *Cold Spring Harb. Perspect. Biol.* 7(2):a016592

144. Renaut S, Rieseberg LH. 2015. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Mol. Biol. Evol.* 32(9):2273–83
145. Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J. 2016. Gene fractionation and function in the ancient subgenomes of maize. *Mol. Biol. Evol.* 34(8):1825–32
146. Riedelsheimer C, Melchinger AE. 2013. Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theor. Appl. Genet.* 126(11):2835–48
147. Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* 273(5281):1516–17
148. Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, et al. 2015. Recombination in diverse maize is stable, predictable, and associated with genetic load. *PNAS* 112(12):3823–28
149. Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. 2016. Open chromatin reveals the functional maize genome. *PNAS* 113(22):E3177–84
150. Saade S, Maurer A, Shahid M, Oakey H, Schmöckel SM, et al. 2016. Yield-related salinity tolerance traits identified in a nested association mapping (NAM) population of wild barley. *Sci. Rep.* 6:32586
151. Sadras VO, Lawson C, Montoro A. 2012. Photosynthetic traits in Australian wheat varieties released between 1958 and 2007. *Field Crops. Res.* 134:19–29
152. Sallam A, Martsch R. 2015. Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica* 143(4):501–14
153. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, et al. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *PNAS* 104(27):11376–81
154. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20(1):43–45
155. Sannemann W, Huang BE, Mathew B, Léon J. 2015. Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Mol. Breed.* 35:86
156. Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *PNAS* 108(10):4069–74
157. Schnable PS, Springer NM. 2013. Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* 64:71–88
158. Schön CC, Dhillon BS, Utz HF, Melchinger AE. 2009. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. *Theor. Appl. Genet.* 120(2):321–32
159. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *PNAS* 111(52):E5661–69
160. Sexton JP, Hangartner SB, Hoffmann AA. 2014. Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* 68(1):1–15
161. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43(2):109–16
162. Soman C, Li D, Wander MM, Kent AD. 2016. Long-term fertilizer and crop-rotation treatments differentially affect soil bacterial community structure. *Plant Soil* 413(1–2):145–59
163. Song WY, Wang GL, Chen LL, Kim HS, Pi LY, et al. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* 270(5243):1804–6
164. Stanton-Geddes J, Paape T, Epstein B, Briskine R, Yoder J, et al. 2013. Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLOS ONE* 8(5):e65688
165. Studer AJ, Doebley JF. 2011. Do large effect QTL fractionate? A case study at the maize domestication QTL *teosinte branched1*. *Genetics* 188(3):673–81
166. Sukumaran S, Reynolds MP, Lopes MS, Crossa J. 2015. Genome-wide association study for adaptation to agronomic plant density: a component of high yield potential in spring wheat. *Crop. Sci.* 55:2609–19
167. Taketa S, Amano S, Tsujino Y, Sato T, Saisho D, et al. 2008. Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *PNAS* 105(10):4062–67
168. Tan B-C, Guan J-C, Ding S, Wu S, Saunders JW, et al. 2017. Structure and origin of the *White Cap* locus and its role in evolution of grain color in maize. *Genetics* 206(1):135–50

169. Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE. 2014. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197(4):1343–55
170. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, et al. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551:457–63
171. Tilman D, Clark M. 2014. Global diets link environmental sustainability and human health. *Nature* 515(7528):518–22
172. Tracy WF, Whitt SR, Buckler ES. 2006. Recurrent mutation and genome evolution: example of and the origin of sweet maize. *Crop. Sci.* 46(Suppl. 1):S49–54
173. Tschermak E. 1900. Ueber künstsliche Kreuzung bei Pisum sativum. In *Berichte der Deutsche Botanischen Gessellschaft*, pp. 232–39. Berlin: Gebrüder Borntræger
174. Turner TR, Ramakrishnan K, Walshaw J, Heavens D, Alston M, et al. 2013. Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *ISME J.* 7(12):2248–58
175. Upadhyaya HD, Vetriventhan M, Deshpande SP, Sivasubramani S, Wallace JG, et al. 2015. Population genetics and structure of a global foxtail millet germplasm collection. *Plant Genome* 8:3
176. Varshney RK, Roorkiwal M, Sorrells ME, eds. 2017. *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*. Cham, Switz.: Springer
177. Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, et al. 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* 35(10):969–76
178. Venuprasad R, Bool ME, Quiatchon L, Sta Cruz MT, Amante M, Atlin GN. 2012. A large-effect QTL for rice grain yield under upland drought stress on chromosome 1. *Mol. Breed.* 30(1):535–47
179. Walker V, Bertrand C, Bellvert F, Moënne-Loccoz Y, Bally R, Comte G. 2011. Host plant secondary metabolite profiling shows a complex, strain-dependent response of maize to plant growth-promoting rhizobacteria of the genus *Azospirillum*. *New Phytol.* 189(2):494–506
180. Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES. 2014. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLOS Genet.* 10(12):e1004845
181. Wallace JG, Upadhyaya HD, Vetriventhan M, Buckler ES, Tom Hash C, Ramu P. 2015. The genetic makeup of a global barnyard millet germplasm collection. *Plant Genome* 8:1
182. Walley JW, Sartor RC, Shen Z, Schmitz RJ, Wu KJ, et al. 2016. Integration of omic networks in a developmental atlas of maize. *Science* 353(6301):814–18
183. Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, et al. 2005. The origin of the naked grains of maize. *Nature* 436(7051):714–19
184. Wang Z, Liu L, Chen Q, Wen X, Liao Y. 2016. Conservation tillage increases soil bacterial diversity in the dryland of northern China. *Agron. Sustain. Dev.* 36(2):28
185. Weber B, Zicola J, Oka R, Stam M. 2016. Plant enhancers: a call for discovery. *Trends Plant Sci.* 21(11):974–87
186. Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102(11):1753–56
187. World Bank. 2016. World development indicators 2016. *World Bank Group*. **http://documents.worldbank.org/curated/en/805371467990952829/World-development-indicators-2016**
188. Wright S. 1921. Systems of mating. I. The biometric relations between parent and offspring. *Genetics* 6(2):111–23
189. Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, et al. 2006. *Sub*1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442(7103):705–8
190. Xu X, Liu X, Ge S, Jensen JD, Hu F, et al. 2011. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30(1):105–11
191. Yahiaoui N, Srichumpa P, Dudler R, Keller B. 2004. Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. *Plant J.* 37(4):528–38
192. Yan J, Kandianis CB, Harjes CE, Bai L, Kim E-H, et al. 2010. Rare genetic variation at *Zea mays crtRB1* increases β-carotene in maize grain. *Nat. Genet.* 42(4):322–27
193. Yang J, Mezmouk S, Baumgarten A, Buckler ES, Guill KE, et al. 2017. Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. *PLOS Genet.* 13(9):e1007019
194. Yao H, Dogra Gray A, Auger DL, Birchler JA. 2013. Genomic dosage effects on heterosis in triploid maize. *PNAS* 110(7):2665–69

195. Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* 353(6306):1431–33
196. York LM, Galindo-Castañeda T, Schussler JR, Lynch JP. 2015. Evolution of US maize (*Zea mays* L.) root architectural and anatomical phenes over the past 100 years corresponds to increased tolerance of nitrogen stress. *J. Exp. Bot.* 66(8):2347–58
197. York LM, Lynch JP. 2015. Intensive field phenotyping of maize (*Zea mays* L.) root crowns identifies phenes and phene integration associated with plant growth and nitrogen acquisition. *J. Exp. Bot.* 66(18):5493–505
198. Zabel F, Putzenlechner B, Mauser W. 2014. Global agricultural land resources–a high resolution suitability evaluation and its perspectives until 2100 under climate change conditions. *PLOS ONE* 9(9):e107522
199. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, et al. 2012. High-resolution mapping of open chromatin in the rice genome. *Genome Res.* 22(1):151–62
200. Zhang Y, Liu Z, Khan AA, Lin Q, Han Y, et al. 2016. Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*Triticum aestivum* L.). *Sci. Rep.* 6:21476
201. Zhao Y, Li Z, Liu G, Jiang Y, Maurer HP, et al. 2015. Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *PNAS* 112(51):15624–29
202. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, et al. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33(4):408–14
203. Zhu X-G, Lynch JP, LeBauer DS, Millar AJ, Stitt M, Long SP. 2016. Plants in silico: Why, why now and what?–an integrative platform for plant systems biology research. *Plant Cell Environ.* 39(5):1049–57