New Strategies and Tools in Quantitative Genetics: How to Go from the Phenotype to the Genotype

Christos Bazakos, Mathieu Hanemian, Charlotte Trontin, José M. Jiménez-Gómez, and Olivier Loudet

Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, 78026 Versailles Cedex, France; email: olivier.loudet@inra.fr

Annu. Rev. Plant Biol. 2017. 68:435–55

First published online as a Review in Advance on February 6, 2017

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-042916-040820

Keywords

genetic architecture, QTL, RIL, GWAS, phenotype

Abstract

Quantitative genetics has a long history in plants: It has been used to study specific biological processes, identify the factors important for trait evolution, and breed new crop varieties. These classical approaches to quantitative trait locus mapping have naturally improved with technology. In this review, we show how quantitative genetics has evolved recently in plants and how new developments in phenotyping, population generation, sequencing, gene manipulation, and statistics are rejuvenating both the classical linkage mapping approaches (for example, through nested association mapping) as well as the more recently developed genome-wide association studies. These strategies are complementary in most instances, and indeed, one is often used to confirm the results of the other. Despite significant advances, an emerging trend is that the outcome and efficiency of the different approaches depend greatly on the genetic architecture of the trait in the genetic material under study.

Contents

1. INTRODUCTION

One main aim in plant genetics is to identify the genes responsible for phenotypic variation associated with adaptive or agronomic traits. Investigating natural variation not only highlights the diversity present, which can be of potential use in crop improvement, but also can provide insights into the evolution and genetic regulation of complex traits. This review focuses on the current developments in phenomics and quantitative genetics strategies and how they are evolving (or will likely evolve) with the advent of modern tools. We focus in particular on some limiting factors and steps in quantitative genetics, such as phenotyping under controlled conditions, genetic material, and mapping strategies. Although we discuss the strong impact of massive sequencing in quantitative genetics, we do not specifically cover sequencing and genotyping methods.

2. PHENOTYPING AT DIFFERENT SCALES

The extensive study and characterization of the quantitative phenotypic variation that a plant genotype can occupy, which corresponds to the theoretical entity of the plant phenome, are key to understanding biological determinants and how the genetic background of a plant interacts with the environment. Phenomics—the comprehensive study of phenotypes—was therefore developed to provide insight into the complex, multidimensional nature of phenotypes (41). However, the lack of phenotyping accuracy in multiple environments and across different spatial and temporal scales is still a systematic limitation for quantitative genetics.

Quantitative genetics: a collection of approaches designed to reveal the genetic architecture of traits of quantitative variation—essentially, linkage mapping in custom-made segregating populations and association mapping among natural populations

In the context of quantitative genetics approaches, phenotyping systems are characterized by dimensionality, resolution, and throughput, and they should meet many (if not all) of the key requirements of such systems, including reproducibility, high throughput, flexibility in environmental treatments, noninvasiveness, and integration of phenotypes at different scales. Reproducibility, which is also a result of homogeneity in both growth conditions and measurement, is one of the most significant factors, and the nongenetic variation caused by uncontrolled environmental perturbations should be avoided or minimized. Throughput is the limiting factor not only for nonautomated phenotyping approaches for large-scale studies, but also for measuring traits that require efficient and frequent screening. Moreover, studying a wide range of traits and conditions requires environmental treatments and monitoring that are diverse enough for use in a range of conditions, from field sites (not specifically covered here) to fully controlled environments. In plant phenotyping platforms, spatial and temporal resolutions are often technically imposed; the spatial scale can range from a single cell to a whole plant, and the temporal resolution can range from seconds to weeks (21).

Another emerging limiting factor in high-throughput phenomics is transferability and scalability, because many of the robotic systems (some of which are detailed below) cannot be easily established by scientists in a new laboratory. By contrast, the popularization of open-source electronics prototyping platforms, such as Arduino and the Raspberry Pi, has allowed users to build simple customized automation modules [for instance, in a simple root phenotyping system (62)] and networks of cheap imaging systems or sensors [for instance, to monitor soil water content and vegetation parameters (9)].

Quantitative trait locus (QTL): a locus that controls the variation of a quantitative trait

Nondestructive phenotyping allows continuous and repeated measurements throughout a plant's life cycle at different scales, from a single cell to a whole root or plant. Beyond imaging in the visible spectrum, image acquisition tools such as near-infrared, fluorescence, infrared (thermal), multi/hyperspectral, X-ray, and magnetic resonance imaging (MRI) technologies can provide a massive amount of phenotyping information from individuals throughout their life spans. These data allow in-depth analysis and integration of phenotypes at different scales in high-throughput experiments, but calibration, signal treatment, and image analysis can remain limiting factors. These phenotyping tools are also valuable for destructive "-omics" studies, including genomics, transcriptomics, proteomics, and metabolomics studies, which are advantageously performed on samples from individuals deeply characterized under precise conditions [for instance, in expression quantitative trait locus (eQTL) studies (20)].

The aim of this review is not to list extensively developed phenotyping systems, although we do mention some typical automated platforms with special features and discuss their potential for quantitative genetic studies. Several of the first phenotyping platforms were devoted to the noninvasive high-throughput imaging of *Arabidopsis thaliana* plants exposed to a variety of biotic and abiotic stresses. Phenopsis enables the automated screening of hundreds of *Arabidopsis* plants for the identification of trait-associated genomic regions (34). Phenoscope was recently developed to minimize the contribution of uncontrolled environmental perturbations despite an increased number of individuals (122). This platform has the particular feature of continuously rotating the pots on the table, exposing all plants to similar micro-environmental variation and greatly improving the homogeneity and reproducibility of quantitative trait locus (QTL) mapping experiments in multiple environments (122).

The noninvasive exploration of root quantitative traits is challenging. GROWSCREEN-Rhizo enables the simultaneous automated measurement of shoot and root growth of small to medium-size plants under various watering and nutrient conditions (84). The main advantage of this methodology results from its characterization of two-dimensional root geometry in correlation with the dynamic plant growth responses to environmental stresses. Recently, three-dimensional plant root imaging has been achieved in soil-growing plants using MRI (125) or X-ray micro-computed tomography (μCT) (81). Both MRI combined with positron emission tomography and X-ray μCT have been successfully applied to the time-series analysis of root-root interactions, providing further insights into the challenges and mechanisms of interacting root systems (24, 74). Although these technologies have provided great insights in the plant root research field, applying them to a wide range of plant root systems still comes with major technical and scientific challenges: Some plant species have a complex root architecture and very thin roots, which in many cases are not detected efficiently. The Growth and Luminescence Observatory for Roots (GLO-Roots) platform uses luminescence-based reporter plants to allow for the visualization of growing roots in soil-filled transparent pots (101). The luminescence enhances the contrast between the root and the soil and thus enables the phenotyping of root architecture and growth regardless of how thin the root is.

Phenotyping facilities are often adapted to a certain plant species, stage, size, or architecture and are not directly scalable for use with other types of plants. Many fully controlled platforms are

dedicated to the study of small rosette plants or the main cereal species (65, 85, 116). Moreover, the spatial resolution of these phenotyping systems is limited to external phenotypes of whole plants or plant organs, without taking into account the internal physiological or structural phenotype (36). Several efforts have been made toward phenotyping at scales down to the cell level by developing tools and computational methods to fuse images acquired from different angles (26, 68). These techniques are difficult to automate for large-scale experiments because they often require essentially manual image acquisition and analysis. Some methods have been used in quantitative genetics studies for cell growth parameters (77, 121).

Linkage mapping: mapping of QTLs in artificial segregating populations

Genetic architecture: the parameters describing the loci that control a trait's segregation; the same trait can have a "simple" genetic architecture in one population and a "complex" architecture in another

To bridge the gap between genotype and external phenotype for complex quantitative traits, the next generation of phenotyping systems should integrate a multidimensional physiological phenotyping that will enable dissection of complex traits into individual physiological components that can be more readily quantified and studied genetically. This gap in knowledge of internal phenotypes, which is far from being filled, is shown clearly in efforts to breed drought-tolerant crops (123).

The recent development of imaging systems has enabled phenotyping platforms to be equipped with a large range of high-resolution sensors for in-depth, high-dimensional phenotyping. Phenovator is an example of a recently developed platform that enables the active and repeated phenotyping of rosette plants for photosynthesis and important leaf pigments (e.g., carotenoids and chlorophyll) throughout plant growth (28). Additionally, the Dynamic Environmental Photosynthetic Imaging (DEPI) platform not only can measure photosynthesis by optical imaging, but also can mimic fluctuating environmental conditions by programmed changes in the room lighting, allowing the identification of novel phenotypes not revealed in standard laboratory conditions (19). Similarly, the SpectralPhenoClimatron growth chamber was developed to provide diurnal and seasonal control of temperature, light color and intensity, and humidity, aiming to fill the gap between controlled conditions and the field in the study of the genetics of adaptation (12, 66).

Other imaging systems are obvious additions to many platforms. Thermal imaging, for example, has been successfully applied for the estimation of transpiration and stomatal conductance to assess drought and heat stress responses. Hyperspectral imaging offers narrow-band data across a wide range of wavelengths that can be correlated with specific physiological features to help with decomposing the QTLs of integrative traits. However, noninvasive phenotyping using fluorescence, thermography, and reflectance can only indirectly assess physiological processes, and it is essential that the results be validated and calibrated with complementary physiological measurements and functional approaches under various environmental conditions and with different species and genotypes. The combination of high-throughput external phenotyping and higher-resolution internal phenotyping will permit more dynamic insights into the plant phenome and link regulatory processes at the molecular level to external plant phenotypes. Last, but not least, the integration and comparison of multiple platforms' phenotypes represent another challenge.

Once phenotypes are obtained for a set of genotypes, linkage mapping and association mapping are commonly used to dissect the genetic architecture of complex traits. In the next sections, we focus on the recent evolution of the material and methods underlying these strategies and the complementarity between them.

3. CLASSICAL QUANTITATIVE GENETICS IS STILL IMPROVING

Linkage mapping analysis in experimental segregating populations is commonly used to dissect the genetic architecture of complex traits. Many different types of segregating populations, derived from the crossing of a limited number of genotypes, are available in plants (**Figure 1**). Starting with the simplest segregating populations, where only two parental accessions are used, an F2

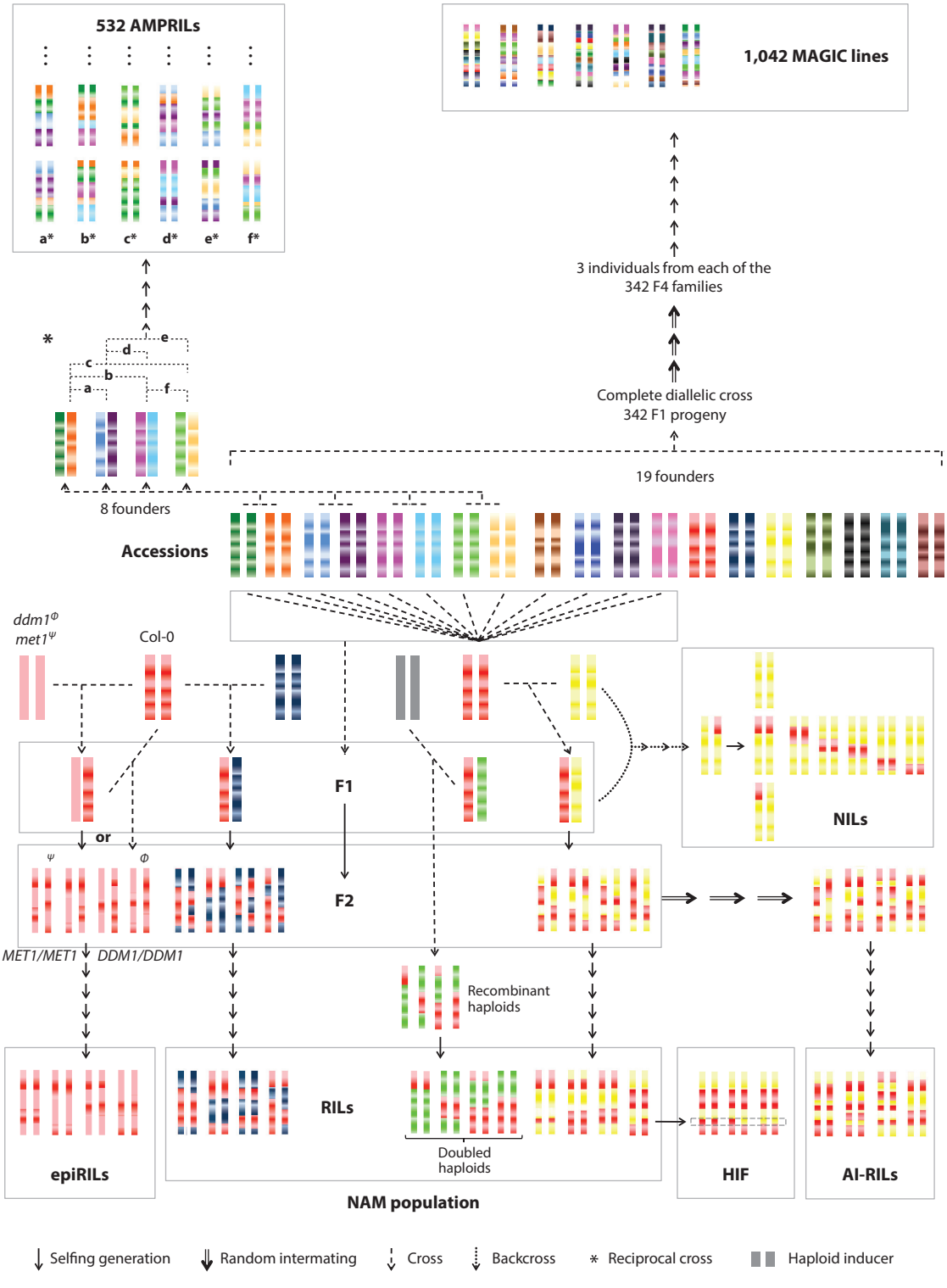
population is the fastest and easiest to generate in most plant species. For example, Salome et al. (110) validated the potential of this approach at a large scale by generating 17 F2 populations from 18 distinct accessions and identifying multiple novel QTLs for flowering time. The many heterozygous regions in F2 populations increase the number of lines required to obtain strong power for additive QTL detection but also allow dominance effects to be tested. Recent advances in sequencing technologies now permit the fast and cost-efficient genotyping of F2 populations and investigation of fundamental processes in plant genetics, such as meiotic recombination (104). However, the need to genotype individuals in each experiment and the impossibility of using the same set of lines in different conditions has pushed plant geneticists to produce recombinant inbred line (RIL) populations from F2 lines by repeated selfing. Each RIL is characterized by high homozygosity and therefore represents a unique mosaic of the two parental genomes (100). RIL populations can be phenotyped repeatedly for different conditions and traits but are genotyped only once because of their homozygosity.

Recombinant inbred line (RIL): a type of segregating population in which two or more parental genomes are shuffled randomly in every line and immortalized at the homozygous state

The long generation time (6–8 generations) of RIL populations was shortened by the creation of doubled haploid (DH) lines that enabled the development of such populations, especially in many crop species, such as barley (140), rice (25), wheat (38, 136), rapeseed (130), and maize (29, 93). DHs are produced by converting haploid plants derived from F1s into diploids, and although this results in fewer recombinations than a RIL population has, it is the fastest way to obtain recombinant homozygous plants. One of the limitations of producing DHs is the requirement for haploid induction, usually through in vitro techniques or rare interspecific crosses, both of which are inhibited in most genotypes (22). To overcome these technical difficulties, Seymour et al. (113) developed for the first time a method to create DHs in *Arabidopsis*. Their recombinant DH population was generated through centromere-mediated genome elimination, taking advantage of the transgenic line *cenh3-1-GFP* (*tailswap*), which contains a modified centromere-specific histone CENH3 protein (99). F1 wild-type plants were crossed to *tailswap* plants to produce recombinant haploids, and DH lines are generated by chromosomal doubling and validated by mapping QTLs for flowering time and petiole length (113).

Once a QTL has been identified, an efficient way to fine-map the contributing loci is to use introgression lines to accumulate extra recombination events while mendelizing the relevant loci. Introgression lines are developed by backcrossing a segregating line with a recurrent genotype until the entire genome except the locus of interest is similar to that of the recurrent parent. Near-isogenic lines (NILs) contain a relatively small homozygous introgressed fragment from one parent in the isogenic background of the other parent (54). The size of the introgressions depends largely on the number of backcross generations performed to obtain the NIL population (8). During construction of these populations, lines can be either genotypically or phenotypically supervised to present the traits or chromosomal regions of interest (78). As with RILs, the permanent homozygous genetic background offers the potential to study traits of interest in various environments and at various time points. NILs have been successfully generated for QTL mapping in many plant species, including *Arabidopsis*, tomato, rice, maize, wheat, and barley. In addition to the use of classical introgression lines, the segregation of a QTL can be confirmed and fine-mapped in heterogeneous inbred families. Such families are developed through the fixation of the heterozygous region in a given F6 or F7 RIL and thus allow the direct comparison of plants with alternative genotypes at the locus of interest in a common (but heterogeneous) homozygous background (69, 124).

Chromosome substitution lines are lines that contain a single chromosome from a different donor genotype. They are valuable for mapping or fine-mapping, either in themselves or as a starting point for NILs, and are typically created through laborious and time-consuming crossing and genotyping during many generations until a specific chromosome substitution line has been


532 AMPRILs
a*
b*
c*
d*
e*
f*
1,042 MAGIC lines
3 individuals from each of the 342 F4 families
Complete diallelic cross 342 F1 progeny
19 founders
8 founders
Accessions
ddm1Φ met1Ψ
Col-0
F1
F2
or
NILs
MET1/MET1
DDM1/DDM1
Recombinant haploids
RILs
Doubled haploids
epiRILs
HIF
AI-RILs
NAM population
Selfing generation
Random intermating
Cross
Backcross
Reciprocal cross
Haploid inducer

identified (59, 109). Wijnen & Keurentjes (131) suggested combining the suppression of recombination in achiasmatic lines [by *DISRUPTION OF MEIOTIC CONTROL 1* (*DMC1*) silencing, as used in the reverse breeding strategy (132)] and *tailswap* haploid-inducer generation to easily create chromosome substitution lines. This opens up a way to generate complex and complete sets of chromosome substitution lines (and, consequently, NILs) in which multiple recurrent backgrounds and introgressed genotypes are used to reveal epistasis and precisely compare allelic series.

The main advantage of classical linkage mapping relies on the power to detect rare variants, significantly increasing the possibility of identifying new alleles of genes involved in phenotypic variation. However, one of the main drawbacks of this approach is the low resolution of QTL confidence intervals (resulting from the limited number of recombination events exploited), which may contain hundreds of genes, resulting in time-consuming and laborious fine-mapping to find the causal gene and polymorphisms (8). Advanced intercross RILs have indeed been generated, by intercrossing several F2-derived generations before fixation, to increase the number of recombination events exploited in RILs (4), but the density of recombination remains limiting.

Bulk segregant analysis (BSA) is another method that offers a high ratio of observed recombination to genotyping effort. Two groups (bulks) of individuals from a segregating population (such as an F2) are selected to display a distinct phenotype distribution in order to skew the allele distribution in each bulk near the genetic interval(s) controlling the phenotype; the allele frequencies of unrelated loci, by contrast, should be equally distributed (32, 80). Takagi et al. (119) showed that BSA is as efficient as classical linkage mapping for mapping QTLs of resistance to the fungal pathogen *Magnaporthe oryzae* and seedling vigor in rice using either F2 or RIL populations. The genotyping of one or two bulks instead of all the individuals of a population is the great advantage of BSA, although of course these bulks are valid only for the specific phenotype used to make the bulks.

The advent of high-throughput genotyping technologies boosted the use of BSA as an efficient method to rapidly identify mutated genes in classical genetic screens (also called mapping by sequencing) (112) as well as to map quantitative traits in segregating populations (also called QTL-seq or extreme QTL mapping) (111, 119, 133) and enabled the application of BSA to many plant species, including tomato (49), *Primula veris* (88), and *Mimulus lewisii* (108). Moreover, BSA can be combined with many sequencing approaches, such as restriction-site-associated DNA sequencing (RAD-seq) or RNA sequencing (RNA-seq), with each combination carrying its own advantages and disadvantages (111). For example, the use of RNA-seq data can be advantageous in species that lack a reference genome and/or have a large genome (117); however, information about intergenic regions and nonexpressed genes will be missing, and allele-specific expression may be misleading regarding allele frequencies if the regulatory sequence is not in linkage disequilibrium (LD) with the called single-nucleotide polymorphisms (SNPs).

Figure 1

Segregating populations for mapping quantitative traits, using *Arabidopsis thaliana* as an example. Each individual's genome is schematically represented by only one chromosome pair. The different colors represent genetic diversity among accessions; darker regions represent methylated alleles, and lighter ones represent unmethylated alleles. This scheme has been simplified to allow the representation of all mapping populations; AMPRIL and MAGIC lines actually have only two founders in common. Abbreviations: AI-RIL, advanced intercross recombinant inbred line; AMPRIL, *Arabidopsis* multiparent recombinant inbred line; Col-0, Columbia 0; *DDM1*, *DECREASED DNA METHYLATION 1*; epiRIL, epigenetic recombinant inbred line (a specific form of induced variation not discussed elsewhere in this review); HIF, heterogeneous inbred family; MAGIC, multiparent advanced generation intercross; *MET1*, *METHYLTRANSFERASE 1*; NAM, nested association mapping; NIL, near-isogenic line.

Heritability: the fraction of an observed trait's variation that is under genetic control

Nested association mapping (NAM): a strategy that uses connected RIL sets to combine some of the advantages of linkage and association mapping

The mapping resolution of QTL-seq can be high if phenotyping is not a bottleneck and a large number of plants can be bulked, as in the study by Rishmawi et al. (102), which led to the direct identification of a candidate gene for a major root hair branching locus. Although the number of individuals per bulk is important for QTL resolution, it is noteworthy that the power of BSA also depends on the sequencing depth and the heritability of the trait studied (23, 37, 73, 111). It is possible to dissect traits in populations with a more complex genetic architecture and to identify major and minor QTLs using this approach, as illustrated by a study of cold tolerance in rice (138) and a study of germination speed in *Arabidopsis* (143).

Overall, BSA is an efficient and cost-effective way to map QTLs and, in the best case, to find candidate genes by overcoming the time-consuming and laborious classical fine-mapping strategy. However, it requires a large plant population and may be more suitable for studying traits that are easily phenotyped and have a high heritability and/or simple genetic architecture in the cross under study. Furthermore, because genotyped individuals are mixed together, BSA cannot directly identify epistatic interactions.

A major drawback of biparental populations is the limited amount of genetic diversity, which by definition is restricted to the contributions of the two founder parents. This issue, in combination with the advance of high-throughput genotyping technologies, motivated interest in generating complex multiparental populations. Although the number of founders is still limited, these populations are characterized by the segregation of much higher allelic diversity. Huang et al. (45) created the *Arabidopsis* multiparent recombinant inbred line (AMPRIL) population by crossing eight *Arabidopsis* founder accessions to generate six independent RIL sets, each consisting of a mosaic of four parental accessions. This population is a valuable resource for studying the allelic diversity of complex traits with great mapping resolution.

An alternative multiparental population scheme that increases the precision of QTL mapping by shuffling the genome and promoting intercrossing is called the multiparent advanced generation intercross (MAGIC) population. Multiple inbred founders are randomly intercrossed over several generations, and the resulting progeny are selfed until homozygosity is reached, thus facilitating the study of quantitative traits and their interaction with the environment. MAGIC populations were first created in *Arabidopsis* (60) and more recently have been created in many crop species, including tomato (90), wheat (42, 72), maize (39), rice (5), and chickpea (31), because the complex pedigree structure of MAGIC populations provides novel breeding opportunities. Additionally, in nonmodel plant species for which a reference genome sequence is not available, they offer a valuable resource for developing high-density linkage maps. MAGIC populations have a higher potential for QTL detection through the segregation of higher genetic diversity because the lines are mosaics of contributions from all founders. However, this advantage is possibly limited by epistasis, which is difficult to resolve because of the limited number of lines representing the numerous possible allelic combinations. Therefore, the experimental design and the choice of parental accessions are crucial for generating a multiparental population and should be adjusted depending on the objective of the derived population and the trait's genetic architecture.

Complementary to linkage mapping approaches, nested association mapping (NAM) (141) links molecular variation with phenotypic variation for complex traits in a star-like cross design. This approach combines the advantages of linkage analysis and association mapping and enables high power and high resolution through joint linkage and association analysis (10). We comment further on this strategy in Section 5.

A recent study in yeast demonstrated a novel breakthrough method for genetic mapping that uses the CRISPR/Cas9 system to perform targeted mitotic recombination events in regions of interest (106). This approach takes advantage of the endonuclease Cas9 and a guide RNA to create a chromosomal double-strand break at a specific site during mitosis. This break leads to a loss of

heterozygosity in heterozygous individuals, making them homozygous from the recombination site to the telomere while leaving them unchanged (heterozygous) everywhere else. This method significantly increases the resolution of genetic mapping without the need for multiple cycles of intercrossing, and its application in plants could have a great impact on classical quantitative genetics approaches.

Roux et al. (103) recently shed light on the effects of cytonuclear interactions on fitness-related traits at the intraspecific level in *Arabidopsis*. By substituting cytoplasmic genomes among eight natural strains, they created a unique series of 56 cytolines to assess their cytonuclear interactions in 28 phenotypic traits. Remarkably, their results indicated that the nuclear genome interacts with the cytoplasmic genome to shape natural variation in 23 out of 28 phenotyping traits.

Genome-wide association study (GWAS): an approach that uses accessions directly to detect genotype-phenotype relationships

4. GENOME-WIDE ASSOCIATION STUDIES BECOME MORE THAN COMPLEMENTARY

In contrast to classical linkage mapping, genome-wide association studies (GWASs) are a population-based approach that takes advantage of the long history of recombination events in natural populations to identify small haplotype blocks associated with phenotypes of interest across species-scale diversity. Genotyping has long been the main limitation of this method, but in the last decade, the combination of decreased sequencing costs and improved data processing has promoted the use of the GWAS approach not only in model species (96) but also in crops (43). GWASs have shed light on the genetic architecture of agronomic traits in many plant species, including rice (46), maize (120), finger and foxtail millet (51, 98), sorghum (114), sugarcane (97), spinach (70), and trees (50). However, as major crops for world food security, rice and maize are at the forefront of crop GWASs. In these species, ambitious multitrait analyses have highlighted important agronomic QTLs useful for further breeding programs (46, 47, 127, 137). Here, we describe some of the advantages and possible limitations of the GWAS approach and discuss why it is complementary to segregating population mapping. **Figure 2** compares the use of these quantitative genetics methods in the context of different genetic architectures.

The use of wild populations that diverged thousands of years ago is the first (if not the main) advantage of the GWAS approach over classical linkage mapping. In regard to this timescale, and depending on the reproductive biology of the species, the accumulation of recombination events across generations reduces the extent of LD and thus ensures a finer exploration of the genome, providing that the marker density is adequate. As a consequence, the resolution of the QTLs identified through a GWAS can directly highlight candidate genes. Because intraspecific and interspecific LD can vary dramatically (27), LD assessment in the association panel used is one of the first steps of GWAS establishment. The LD measure should determine the marker density required to ensure the detection of all the haplotypes along the genome. In *Arabidopsis*, LD decays rapidly (within 10 kb, in most cases), so the optimal number of SNPs necessary to indirectly survey any polymorphism in the genome (~135 Mb) has been estimated at between 140,000 and 240,000, and SNP arrays were initially designed accordingly (40, 55). However, LD is not homogeneous along the genome, and even in *Arabidopsis*, an association block may still cover several candidate genes.

With sequencing costs still falling and new methods available to reduce the processing time of the statistical analysis, it is now feasible to use full genomic information in a GWAS. A comparison of the use of resequencing data with the use of a SNP array in *Arabidopsis* showed that a denser genotypic set can still be beneficial and improve statistical power (1). Moreover, whole-genome sequencing allows one to screen all polymorphisms that have a potential impact on gene function in the confidence interval of the association peak and to identify candidate genes. Yano et al. (139)

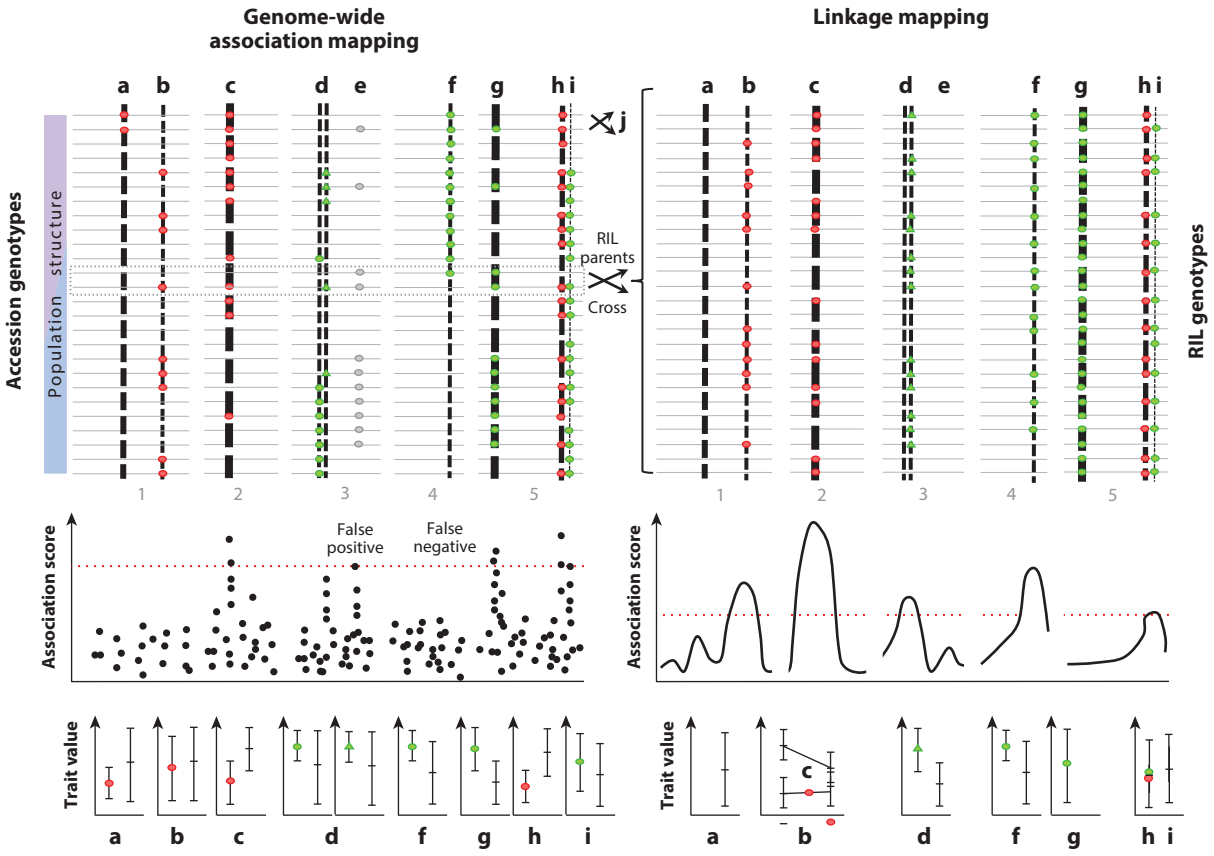


Figure 2

The effect of genetic architecture on the power to detect significant QTLs using linkage and association mapping. The genetic architecture of a theoretical trait is represented in a set of accessions (*left*) and in a population of RILs (*right*) descending from a specific cross (*gray dotted box*). On the five chromosomes of each individual (*gray horizontal lines*), the alleles with positive (*green*) or negative (*red*) effects on the trait are indicated along vertical dashed bars (representing the QTL position), with the thickness of each bar representing the strength of the QTL effect. The association score detected for each mapping method (with the significance thresholds indicated as red dotted lines) and the phenotypic trait values associated with each QTL's alleles are also indicated at the bottom of the illustration. Specific theoretical cases: (*a*) Association mapping has poor power to detect rare variants. Such variants will be detected by linkage mapping only if the variant segregates between the chosen parents of the RIL population. (*b*,*c*) Association mapping may not detect some loci because they are in epistatic interaction. Epistatic loci can be detected using special statistical analysis in linkage mapping data when the population is large enough or the interaction strong enough. (*d*) Allelic heterogeneity reduces the power of association mapping. (*e*–*g*) The genetic background effect can cause spurious associations (false positives). Here, polymorphism *e* (not causal) is in linkage disequilibrium with polymorphism *g* (causal) and therefore emerges from the GWAS analysis. After correction for population structure, some loci may no longer be detected (false negatives) because of correlations between QTL allele distribution and population structure (*f*, causal). (*h*,*i*) Compared with the GWAS approach, linkage mapping has poor power to detect linked QTLs, particularly when they have opposite effects. Depending on the alleles segregating in the parents of the RIL population, loci detected using GWAS mapping may or may not be detected by linkage mapping. For example, in the RIL population issued from cross *j*, linkage mapping would detect only one QTL (*g*). Abbreviations: GWAS, genome-wide association study; QTL, quantitative trait locus; RIL, recombinant inbred line.

recently used this approach in a GWAS conducted on rice to rapidly identify both previously known and new candidate genes within the association confidence intervals. Another interesting point in this study was the characterization of a causal gene located more than 1 Mb away from the closest association peak as a result of allelic heterogeneity.

LD extent is often related to the species mating system. In this regard, outcrossing allows more mixing of genetic diversity between individuals of a population than does selfing, which can lead to almost pure homozygous lines (71). For example, in rice, LD in the autogamous cultivated subspecies *Oryza sativa* decays over ~150 kb, whereas in lines of the wild outcrossing species *Oryza rufipogon*, it decays rapidly over ~20 kb (46, 75). Huang et al. (44) therefore estimated that the resolution of the GWAS mapping in *O. rufipogon* is three times better than that in *O. sativa*. The reduction of diversity that typically accompanies domestication can also affect the size of haplotype blocks, as exemplified by the difference between cultivated and wild barley (both of which are selfing species), which have LD extents of several hundred kilobases and a few kilobases, respectively (14, 82). In maize, a typical outcrossing species, LD decays within ~2 kb (33, 134); maize is therefore well suited for GWASs, as this approach should point directly to candidate genes. For example, a GWAS conducted on maize seedlings subjected to drought stress identified 42 candidate genes (129). Looking at the most significant SNPs located on chromosome 9 and using further functional and resequencing analysis, the authors demonstrated that a 366-base-pair insertion in its promoter promotes the drought responsiveness of the *Zea mays vacuolar-type H^+ pyrophosphatase 1* (*ZmVPP1*) gene. Meijon et al. (77) carried out an impressive GWAS designed to study root morphology at the cellular level in *Arabidopsis*. Expression analysis of the genes in the confidence interval of the most significant association (albeit below the threshold) led to the identification of a new F-box gene named *KURZ UND KLEIN* (*KUK*), which is involved in the length of the root apical meristem and mature cortical cells.

Overall, these examples illustrate that GWAS resolution can vary dramatically, from the level of individual genes to several hundred kilobases, depending on the LD of the association panel and other parameters (including population structure; see below). In addition, the most significantly associated SNP will not necessarily be the closest to the causal gene or polymorphism (53), and there are likely to be false negatives (2, 16). The amount of genotypic data used for GWASs of hundreds of natural lines provides a better picture of species-wide allelic diversity, frequency, and combination than linkage mapping does. This difference is relevant for studying what happens in nature (or during the domestication/breeding process) and highlights the forces driving natural and/or agronomic selection (79).

However, the broader genetic base of GWASs is counterbalanced by the difficulty of identifying rare alleles, which can represent a large fraction of the genetic variation (47, 63, 105, 135) and are particularly interesting when studying adaptation because many traits seem to be prone to genetic and allelic heterogeneity (6, 48, 107). Rare alleles (e.g., those with a frequency of <5% in the population) are typically not used in GWASs because they are not assessed properly owing to lack of representation in the studied set. As the statistical power of a GWAS depends on both the allelic effect on the phenotype and the allele frequency, there is little power to detect even moderately frequent alleles of small effect (57). In the case of epistasis—another complicating factor in GWASs—sample size is again the primary limitation, but computing power is also a constraint, although methods have emerged to address this challenge (145). From this perspective, a GWAS can be conducted as a first step toward uncovering the genetic architecture of complex traits (128, 146) and toward selecting two or more natural lines from an association panel in order to generate a segregating population for further confirmation of linkage, illustrating the complementarity of association and linkage mapping (10). Working in a well-chosen homozygous

segregating population artificially restores a rare allele to a frequency of ~50% and each allelic combination from a bilocus epistasis to 25%.

RNA-seq can be coupled with GWAS to target genotypic information to expressed genes and subsequently analyze candidate genes' transcription. A GWAS conducted on 368 maize inbred lines genotyped with more than a million expressed SNPs identified 74 loci involved in kernel oil concentration and fatty acid composition (63). The functional annotation of the genes in the confidence intervals identified several candidate genes involved in lipid metabolism. Although RNA-seq data do not provide information about noncoding sequences, it is possible to resequence candidate genes and identify causal mutations. RNA-seq data can also be used to estimate the expression correlation between candidate genes and with other genes and to build complex coexpression networks (63). Moreover, RNA-seq data can be used directly as phenotypes for association in order to study the genetic basis of expression variation (18, 30, 52), an approach that is advantageously complemented by allele-specific expression assays in hybrids (20).

Epigenetic variation can also generate phenotypic diversity. Epigenetic marks usually remain linked to genotypic information in segregating populations (and therefore are taken into account by mapping) but may also be at least partly shuffled from DNA haplotypes in GWAS populations, requiring that these marks be considered independently. Thus, recent work has begun to include epigenetic information in the source of variation (52).

The GWAS limitation that has received the most attention is the confounding effect arising from the genetic background. Population structure, which generally describes the remote common ancestry of large groups of individuals, can cause LD between the causal variants of a given trait and unlinked loci throughout the genome, leading to spurious genotype-phenotype associations (**Figure 2**). This occurrence of false positives is particularly important when the addressed phenotypic variation overlaps with the pattern of population structure and/or environmental clines (126). Phenomena such as allelic incompatibilities, although more limited in scale, may cause LD in the genetic background and result in the same spurious associations. Several methods have been developed to take population structure into account, such as using the cluster membership obtained with STRUCTURE software (95) or using the largest variance-explaining factor detected by principal component analysis (94) as a fixed effect in the association model. An additional improvement, based on Fisher's observation in 1918 that similarity between individuals correlates with the number of shared alleles, models phenotype with a linear mixed model that accounts for the phenotypic variation that is linked with accessions' pairwise relatedness (83, 142). Overall, these methods reduce the number of false positives but also reduce sensitivity (67). The appearance of false negatives when correcting population relatedness is especially likely when the environmental conditions that constrain phenotypic variation and genetic variants overlap with population structure. In addition, these models can artificially inflate the association score of rare alleles, leading to an increase in the minimal minor allele frequency (sometimes up to 10%) (11). Overall, it should be noted that the most direct way to confirm an association and clearly rule out a false positive remains to study the segregation of the locus in a relevant cross by linkage mapping.

Undetected rare alleles and confounding genetic background effects could also be partially avoided by working with GWAS samples from regional scales. This approach decreases the variation in relatedness and the number of genes and alleles per locus that contribute to the phenotypic variation observed for a given trait (11, 57). However, the countereffect is a reduction of the genetic diversity uncovered and possibly an increase in the average LD extent (15), limiting the resolution of GWAS mapping. It is also possible to run GWAS first on the whole association panel and then on subpopulations to see whether new associations arise. For example, a GWAS conducted on an association panel of 367 maize inbred lines led to the identification of 42 significantly associated

SNPs; the use of 152 tropical/subtropical lines and 149 temperate lines independently allowed the authors to identify 21 and 20 new associated SNPs, respectively (129).

The association panel design is critical for GWAS success. A good design requires taking into account the population structure and relatedness as well as optimizing the genotypic and phenotypic diversity in order to find the best balance among those parameters. Core association panels have been developed in several species based on these criteria (56, 87, 114). For instance, from a maize collection of 632 inbred lines, Yan et al. (134) were able to survey 90% of the haplotype diversity with only 60 lines. GWASs can be conducted on a small association panel, most notably for traits with a simple genetic architecture and high heritability. The resistance gene *RESISTANCE TO P. SYRINGAE PV MACULICOLA 1* (*RPM1*) (35) was easily found by a GWAS of fewer than 100 *Arabidopsis* accessions based on the hypersensitive response triggered by the bacterial avirulence gene *AvrRPM1* (2). However, even in the case of a simple genetic architecture, the use of a larger association panel provides stronger statistical power. In studies of sodium concentration in *Arabidopsis* leaves, GWASs performed with association panels of 93 and 349 lines provided 2 and 12 associated SNPs, respectively, within *HIGH-AFFINITY* K^+ *TRANSPORTER 1* (*HKT1*), a causal gene that encodes a sodium transporter (2, 7). In a study of cadmium concentration in *Arabidopsis* leaves, only the association panel of 349 lines enabled the detection of *HEAVY METAL ATPASE 3* (*HMA3*), a causal gene that encodes a major component of cadmium homeostasis in *Arabidopsis* (17).

Finally, multitrait and/or multienvironment models for GWASs using mixed models can increase the power to detect associations (58). At least in certain species, the GWAS approach is becoming the strategy of choice for quantitative genetics, especially when the development of more advanced genetic material is difficult.

5. IS NESTED ASSOCIATION MAPPING THE RIGHT BALANCE?

NAM is a recently developed strategy for finely mapping QTLs that aims to combine the power of QTL detection in linkage analyses and the resolution of association mapping approaches (141). NAM was initially developed in maize by Buckler et al. (13), who used a multiparental mapping population that consisted of 25 founder lines crossed with one common line in a star-like crossing design, resulting in 25 RIL sets connected by one parent. They then genotyped the RILs and parental inbreds with common-parent-specific markers at a relatively low density and inferred the inheritance of the genome region nested within two of these markers through linkage. Common-parent-specific alleles are used to normalize the genetic background for mapping of segregating alleles in founder lines and to relativize the effect of genetic heterogeneity. Joint linkage mapping is performed on the phenotypic data of all RILs simultaneously in a single model, taking into account the RIL family effect while providing an estimate of the total variance associated with a genomic region. This strategy exploits both ancient and recent recombination events in a supposedly balanced way and increases mapping resolution and power by effectively combining the strengths of both association and linkage mapping (141). Ogut et al. (89) recently used a maize NAM population to compare joint-family and single-family QTL mapping in order to evaluate their relative power to detect rare small-effect QTLs. Their results suggested that, although joint-family QTL models still miss some rare large-effect QTLs, they have the power to detect a larger number of small-effect QTLs. Hence, joint-family and single-family analyses can be complementary.

After intense material development and sharing, NAM has been widely used in maize to study various traits, including flowering time (13), disease resistance (61, 92), plant architecture (91), and carbon and nitrogen metabolism (144). Recent studies have also exploited the advantages of

NAM populations in barley (76) and wheat (3). In *Arabidopsis*, multiparent populations such as MAGIC lines and AMPRILs share some of NAM's advantages (see Section 3), but no approach specifically uses NAM statistics yet. However, there are several connected RIL sets in *Arabidopsis* that should allow the use of this approach (115).

Drawbacks of NAM result from the need to carefully choose the crossing design based on the number of founders, the need for large-scale phenotyping of thousands of plants, and the presence of genetic and allelic heterogeneity (8). As discussed in Section 3, the development of RILs is labor intensive and requires several generations. New mating designs in a NAM context, such as doubled haploid NAM (DH-NAM), backcross NAM (BC-NAM) (64), and advanced backcross NAM (AB-NAM) (86), have been suggested in order to shorten this time. NAM is a promising strategy that can offer the right balance between linkage analysis and association mapping. However, each mating design should be adjusted to the species, taking into consideration the genetic diversity and reproductive system of the plants (i.e., whether they allow easy crossing and selfing) and keeping in mind that the number of parental inbreds should be larger for species with a large number of rare alleles (118).

6. FINAL REMARK

Overall, the decisive limiting factor common to all quantitative genetics approaches is the genetic architecture of the trait under consideration in the studied population, which remains difficult to predict before the mapping has been performed. Among the features that will be decisive for the efficiency of the approach are the number and distribution of loci, number of alleles per loci, explained variance per loci, additivity/dominance, epistatic interaction between loci, frequency of the allele in the population, and interactions with the environment. Preliminary exploration of these factors in subpopulations in order to estimate a trait's heritability and complexity under specific phenotyping conditions can help investigators avoid disappointment.

SUMMARY POINTS

1. The genetic architecture of a trait within the genetic material is the key factor that determines the efficiency of different quantitative genetics strategies.
2. Recently emerged tools such as haploid-inducer and achiasmatic lines allow advanced material (such as recombinant inbred lines, near-isogenic lines, and chromosome substitution lines) to be generated more easily.
3. Although genome-wide association approaches require an initial investment, they have the advantage of high mapping resolution. Their outcome depends heavily on the specific trait's genetic architecture in the chosen population.
4. Nested association mapping, which uses new methods of statistical analysis in connected sets of traditional recombinant inbred lines, combines many advantages of linkage and association mapping.

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

This work was supported by funding from the European Commission Framework Programme 7 ERC Starting Grant "DECODE" (ERC-2009-StG-243359) to O.L. and ANR grant "2Complex" (ANR-09-BLAN-0366) to O.L. The Institut Jean-Pierre Bourgin benefits from the support of the LabEx Saclay Plant Sciences (ANR-10-LABX-0040-SPS).

LITERATURE CITED

1. **1001 Genomes Consort. 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166:481–91**
2. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, et al. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–31
3. Bajgain P, Rouse MN, Tsilo TJ, Macharia GK, Bhavani S, et al. 2016. Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLOS ONE* 11:e0155760
4. Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, et al. 2009. QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. *PLOS ONE* 4:e4318
5. Bandillo N, Raghavan C, Muyco PA, Sevilla MA, Lobina IT, et al. 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6:11
6. Barboza L, Effgen S, Alonso-Blanco C, Kooke R, Keurentjes JJ, et al. 2013. *Arabidopsis* semidwarfs evolved from independent mutations in *GA20ox1*, ortholog to green revolution dwarf alleles in rice and barley. *PNAS* 110:15818–23
7. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, et al. 2010. A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1. *PLOS Genet.* 6:e1001193
8. Bergelson J, Roux F. 2010. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat. Rev. Genet.* 11:867–79
9. Bitella G, Rossi R, Bochicchio R, Perniola M, Amato M. 2014. A novel low-cost open-hardware platform for monitoring soil water content and multiple soil-air-vegetation parameters. *Sensors* 14:19639–59
10. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, et al. 2010. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLOS Genet.* 6:e1000940
11. Brachi B, Morris GP, Borevitz JO. 2011. Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biol.* 12:232
12. Brown TB, Cheng R, Sirault XR, Rungrat T, Murray KD, et al. 2014. TraitCapture: genomic and environment modelling of plant phenomic data. *Curr. Opin. Plant Biol.* 18:73–79
13. **Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, et al. 2009. The genetic architecture of maize flowering time. *Science* 325:714–18**
14. Caldwell KS, Russell J, Langridge P, Powell W. 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172:557–67
15. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, et al. 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43:956–63
16. Chan EK, Rowe HC, Kliebenstein DJ. 2010. Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185:991–1007
17. Chao DY, Silva A, Baxter I, Huang YS, Nordborg M, et al. 2012. Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in *Arabidopsis thaliana*. *PLOS Genet.* 8:e1002923
18. Clauw P, Coppens F, Korte A, Herman D, Slabbinck B, et al. 2016. Leaf growth response to mild drought: natural variation in Arabidopsis sheds light on trait architecture. *Plant Cell* 28:2417–34
19. Cruz JA, Savage LJ, Zegarac R, Hall CC, Satoh-Cruz M, et al. 2016. Dynamic environmental photosynthetic imaging reveals emergent phenotypes. *Cell Syst.* 2:365–77
20. Cubillos FA, Stegle O, Grondin C, Canut M, Tisne S, et al. 2014. Extensive *cis*-regulatory variation robust to environmental perturbation in *Arabidopsis*. *Plant Cell* 26:4298–310

1. Provides genome sequence data for 1,135 *Arabidopsis* accessions, setting up GWAS at full genomic scale.

13. The founding paper that established and exploited NAM in maize.

21. Dhondt S, Wuyts N, Inzé D. 2013. Cell to whole-plant phenotyping: The best is yet to come. *Trends Plant Sci*. 18:428–39
22. Dunwell JM. 2010. Haploids in flowering plants: origins and exploitation. *Plant Biotechnol. J.* 8:377–424
23. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, et al. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464:1039–42
24. Faget M, Nagel KA, Walter A, Herrera JM, Jahnke S, et al. 2013. Root-root interactions: extending our perspective to be more inclusive of the range of theories in ecology and agriculture using in-vivo analyses. *Ann. Bot.* 112:253–66
25. Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang Q. 2005. The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. *Theor. Appl. Genet.* 110:1445–52
26. Fernandez R, Das P, Mirabet V, Moscardi E, Traas J, et al. 2010. Imaging plant growth in 4D: robust tissue reconstruction and lineaging at cell resolution. *Nat. Methods* 7:547–53
27. Flint-Garcia SA, Thornsberry JM, Buckler ES. 2003. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54:357–74
28. Flood PJ, Kruijer W, Schnabel SK, van der Schoor R, Jalink H, et al. 2016. Phenomics for photosynthesis, growth and reflectance in *Arabidopsis thaliana* reveals circadian and long-term fluctuations in heritability. *Plant Methods* 12:14
29. Foiada F, Westermeier P, Kessel B, Ouzunova M, Wimmer V, et al. 2015. Improving resistance to the European corn borer: a comprehensive study in elite maize using QTL mapping and genome-wide prediction. *Theor. Appl. Genet.* 128:875–91
30. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, et al. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477:419–23
31. Gaur PM, Jukanti AK, Varshney RK. 2012. Impact of genomic technologies on chickpea breeding strategies. *Agronomy* 2:199–221
32. Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Res.* 19:6553–58
33. Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, et al. 2009. A first-generation haplotype map of maize. *Science* 326:1115–17
34. Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, et al. 2006. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.* 169:623–35
35. Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, et al. 1995. Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* 269:843–46
36. Grosskinsky DK, Svensgaard J, Christensen S, Roitsch T. 2015. Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J. Exp. Bot.* 66:5429–40
37. Guo L, Ganguly A, Sun L, Suo F, Du LL, Russell P. 2016. Global fitness profiling identifies arsenic and cadmium tolerance mechanisms in fission yeast. *G3* 6:3317–33
38. Hill CB, Taylor JD, Edwards J, Mather D, Langridge P, et al. 2015. Detection of QTL for metabolic and agronomic traits in wheat with adjustments for variation at genetic loci that affect plant phenology. *Plant Sci.* 233:143–54
39. Holland JB. 2015. MAGIC maize: a new resource for plant genetics. *Genome Biol.* 16:163
40. Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, et al. 2012. Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat. Genet.* 44:212–16
41. Houle D, Govindaraju DR, Omholt S. 2010. Phenomics: the next challenge. *Nat. Rev. Genet.* 11:855–66
42. Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, et al. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol. J.* 10:826–39
43. Huang X, Han B. 2014. Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65:531–51

44. Huang X, Kurata N, Wei X, Wang ZX, Wang A, et al. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:497–501
45. Huang X, Paulo MJ, Boer M, Effgen S, Keizer P, et al. 2011. Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. *PNAS* 108:4488–93
46. Huang X, Wei X, Sang T, Zhao Q, Feng Q, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42:961–67
47. Huang X, Zhao Y, Wei X, Li C, Wang A, et al. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* 44:32–39
48. Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, et al. 2013. An atypical kinase under balancing selection confers broad-spectrum disease resistance in Arabidopsis. *PLOS Genet*. 9:e1003766
49. Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E. 2015. Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theor. Appl. Genet.* 128:1329–42
50. Iwata H, Minamikawa MF, Kajiya-Kanegae H, Ishimori M, Hayashi T. 2016. Genomics-assisted breeding in fruit trees. *Breed. Sci*. 66:100–15
51. Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, et al. 2013. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* 45:957–61
52. Kawakatsu T, Huang SS, Jupe F, Sasaki E, Schmitz RJ, et al. 2016. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* 166:492–505
53. Kerdaffrec E, Filiault DL, Korte A, Sasaki E, Nizhynska V, et al. 2016. Multiple alleles at a single locus control seed dormancy in Swedish *Arabidopsis*. *eLife* 5:e22502
54. Keurentjes JJ, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, et al. 2007. Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 175:891–905
55. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, et al. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 39:1151–55
56. Kim TS, He Q, Kim KW, Yoon MY, Ra WH, et al. 2016. Genome-wide resequencing of KRICE_CORE reveals their potential for future breeding, as well as functional and evolutionary studies in the post-genomic era. *BMC Genom*. 17:408
57. Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29
58. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* 44:1066–71
59. Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, et al. 2002. STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. *Plant J.* 31:355–64
60. Kover PX, Rowntree JK, Scarcelli N, Savriama Y, Eldridge T, Schaal BA. 2009. Pleiotropic effects of environment-specific adaptation in *Arabidopsis thaliana*. *New Phytol.* 183:816–25
61. Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, et al. 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* 43:163–68
62. Le Marié C, Kirchgessner N, Marschall D, Walter A, Hund A. 2014. Rhizoslides: paper-based growth system for non-destructive, high throughput phenotyping of root development by means of image analysis. *Plant Methods* 10:13
63. Li H, Peng Z, Yang X, Wang W, Fu J, et al. 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45:43–50
64. Li J, Bus A, Spamer V, Stich B. 2016. Comparison of statistical models for nested association mapping in rapeseed (*Brassica napus* L.) through computer simulations. *BMC Plant Biol*. 16:26
65. Li L, Zhang Q, Huang D. 2014. A review of imaging techniques for plant phenotyping. *Sensors* 14:20078–111
66. Li Y, Cheng R, Spokas KA, Palmer AA, Borevitz JO. 2014. Genetic variation for life history sensitivity to seasonal warming in *Arabidopsis thaliana*. *Genetics* 196:569–77
67. Lipka AE, Kandianis CB, Hudson ME, Yu J, Drnevich J, et al. 2015. From association to prediction: statistical methods for the dissection and selection of complex traits in plants. *Curr. Opin. Plant Biol.* 24:110–18

52. Provides epigenome variation data for 1,107 *Arabidopsis* accessions and transcriptome data for 1,203 accessions.

68. Liu M, Chakraborty A, Singh D, Yadav RK, Meenakshisundaram G, et al. 2011. Adaptive cell segmentation and tracking for volumetric confocal microscopy images of a developing plant meristem. *Mol. Plant* 4:922–31
69. Loudet O, Gaudon V, Trubuil A, Daniel-Vedele F. 2005. Quantitative trait loci controlling root growth and architecture in *Arabidopsis thaliana* confirmed by heterogeneous inbred family. *Theor. Appl. Genet.* 110:742–53
70. Ma J, Shi A, Mou B, Evans M, Clark JR, et al. 2016. Association mapping of leaf traits in spinach (*Spinacia oleracea* L.). *Plant Breed.* 135:399–404
71. Mable BK, Adam A. 2007. Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. *Mol. Ecol.* 16:3565–80
72. Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, et al. 2014. An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3* 4:1603–10
73. Magwene PM, Willis JH, Kelly JK. 2011. The statistics of bulk segregant analysis using next generation sequencing. *PLOS Comput. Biol.* 7:e1002255
74. Mairhofer S, Sturrock CJ, Bennett MJ, Mooney SJ, Pridmore TP. 2015. Extracting multiple interacting root systems using X-ray microcomputed tomography. *Plant J.* 84:1034–43
75. Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD. 2007. The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177:2223–32
76. Maurer A, Draba V, Pillen K. 2016. Genomic dissection of plant development and its impact on thousand grain weight in barley through nested association mapping. *J. Exp. Bot.* 67:2507–18
77. Meijon M, Satbhai SB, Tsuchimatsu T, Busch W. 2014. Genome-wide association study using cellular traits identifies a new regulator of root development in *Arabidopsis*. *Nat. Genet.* 46:77–81
78. Méndez-Vigo B, Martínez-Zapater JM, Alonso-Blanco C. 2013. The flowering repressor *SVP* underlies a novel *Arabidopsis thaliana* QTL interacting with the genetic background. *PLOS Genet.* 9:e1003289
79. Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, et al. 2016. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* 48:1083–88
80. Michelmore RW, Paran I, Kesseli RV. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *PNAS* 88:9828–32
81. Mooney SJ, Pridmore TP, Helliwell J, Bennett MJ. 2012. Developing X-ray computed tomography to non-invasively image 3-D root systems architecture in soil. *Plant Soil* 352:1–22
82. Morrell PL, Toleno DM, Lundy KE, Clegg MT. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *PNAS* 102:2442–47
83. Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, et al. 2009. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194–202
84. Nagel KA, Putz A, Gilmer F, Heinz K, Fischbach A, et al. 2012. GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. *Funct. Plant Biol.* 39:891–904
85. Neilson EH, Edwards AM, Blomstedt CK, Berger B, Moller BL, Gleadow RM. 2015. Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C_4 cereal crop plant to nitrogen and water deficiency over time. *J. Exp. Bot.* 66:1817–32
86. Nice LM, Steffenson BJ, Brown-Guedira GL, Akhunov ED, Liu C, et al. 2016. Development and genetic characterization of an advanced backcross–nested association mapping (AB-NAM) population of wild × cultivated barley. *Genetics* 203:1453–67
87. Nicolas SD, Peros JP, Lacombe T, Launay A, Le Paslier MC, et al. 2016. Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L) diversity panel newly designed for association studies. *BMC Plant Biol.* 16:74
88. Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E. 2015. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biol.* 16:12
89. Ogut F, Bian Y, Bradbury PJ, Holland JB. 2015. Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. *Heredity* 114:552–63

77. A great example of a successful GWAS to find new genes for cellular traits.

90. 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.* 13:565–77
91. Peiffer JA, Flint-Garcia SA, De Leon N, McMullen MD, Kaeppler SM, Buckler ES. 2013. The genetic architecture of maize stalk strength. *PLOS ONE* 8:e67066
92. Poland JA, Bradbury PJ, Buckler ES, Nelson RJ. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *PNAS* 108:6893–98
93. Presterl T, Ouzunova M, Schmidt W, Moller EM, Rober FK, et al. 2007. Quantitative trait loci for early plant vigour of maize grown in chilly environments. *Theor. Appl. Genet.* 114:1059–70
94. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38:904–9
95. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59
96. Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, et al. 2016. 50 years of Arabidopsis research: highlights and future directions. *New Phytol.* 209:921–44
97. Racedo J, Gutiérrez L, Perera MF, Ostengo S, Pardo EM, et al. 2016. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol.* 16:142
98. Ramakrishnan M, Antony Ceasar S, Duraipandiyan V, Vinod KK, Kalpana K, et al. 2016. Tracing QTLs for leaf blast resistance and agronomic performance of finger millet (*Eleusine coracana* (L.) Gaertn.) genotypes through association mapping and in silico comparative genomics analyses. *PLOS ONE* 11:e0159264
99. Ravi M, Chan SW. 2010. Haploid plants produced by centromere-mediated genome elimination. *Nature* 464:615–18
100. Reiter RS, Williams J, Feldman K, Rafalski JA, Tingey SV, Scolnik PA. 1992. Global and local genome mapping in *Arabidopsis thaliana* recombinant inbred lines and random amplified polymorphic DNAs. *PNAS* 89:1477–81
101. Rellán-Álvarez R, Lobet G, Lindner H, Pradier PL, Sebastian J, et al. 2015. GLO-Roots: an imaging platform enabling multidimensional characterization of soil-grown root systems. *eLife* 4:e07597
102. Rishmawi L, Sun H, Schneeberger K, Hulskamp M, Schrader A. 2014. Rapid identification of a natural knockout allele of *ARMADILLO REPEAT-CONTAINING KINESIN1* that causes root hair branching by mapping-by-sequencing. *Plant Physiol.* 166:1280–87
103. Roux F, Mary-Huard T, Barillot E, Wenes E, Botran L, et al. 2016. Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *PNAS* 113:3687–92
104. Rowan BA, Patel V, Weigel D, Schneeberger K. 2015. Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3* 5:385–98
105. Ruggieri V, Francese G, Sacco A, D'Alessandro A, Rigano MM, et al. 2014. An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biol.* 14:337
106. Sadhu MJ, Bloom JS, Day L, Kruglyak L. 2016. CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science* 352:1113–16
107. Saez-Aguayo S, Rondeau-Mouro C, Macquet A, Kronholm I, Ralet MC, et al. 2014. Local evolution of seed flotation in Arabidopsis. *PLOS Genet.* 10:e1004221
108. Sagawa JM, Stanley LE, LaFountain AM, Frank HA, Liu C, Yuan YW. 2016. An R2R3-MYB transcription factor regulates carotenoid pigmentation in *Mimulus lewisii* flowers. *New Phytol.* 209:1049–57
109. Saha S, Wu J, Jenkins JN, McCarty JC, Hayes R, Stelly DM. 2010. Genetic dissection of chromosome substitution lines of cotton to discover novel *Gossypium barbadense* L. alleles for improvement of agronomic traits. *Theor. Appl. Genet.* 120:1193–205
110. Salome PA, Bomblies K, Laitinen RA, Yant L, Mott R, Weigel D. 2011. Genetic architecture of flowering-time variation in *Arabidopsis thaliana*. *Genetics* 188:421–33
111. Schlotterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* 15:749–63
112. Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, et al. 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* 6:550–51
113. Seymour DK, Filiault DL, Henry IM, Monson-Miller J, Ravi M, et al. 2012. Rapid creation of *Arabidopsis* doubled haploid lines for quantitative trait locus mapping. *PNAS* 109:4227–32

113. Describes the first RIL population fixed through centromere-mediated genome-elimination mutants (haploid inducer).

114. Shakoor N, Ziegler G, Dilkes BP, Brenton Z, Boyles R, et al. 2016. Integration of experiments across diverse environments identifies the genetic determinants of variation in *Sorghum bicolor* seed element composition. *Plant Physiol.* 170:1989–98
115. Simon M, Loudet O, Durand S, Bérard A, Brunel D, et al. 2008. QTL mapping in five new large RIL populations of *Arabidopsis thaliana* genotyped with consensus SNP markers. *Genetics* 178:2253–64
116. Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, et al. 2011. Survival and growth of *Arabidopsis* plants given limited water are not equal. *Nat. Biotechnol.* 29:212–14
117. Sloan DB, Keller SR, Berardi AE, Sanderson BJ, Karpovich JF, Taylor DR. 2012. De novo transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae). *Mol. Ecol. Resour.* 12:333–43
118. Stich B. 2009. Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics* 183:1525–34
119. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, et al. 2013. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74:174–83
120. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, et al. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* 43:159–62
121. Tisné S, Reymond M, Vile D, Fabre J, Dauzat M, et al. 2008. Combined genetic and modeling approaches reveal that epidermal cell area and number in leaves are controlled by leaf and plant developmental processes in Arabidopsis. *Plant Physiol.* 148:1117–27
122. Tisné S, Serrand Y, Bach L, Gilbault E, Ben Ameur R, et al. 2013. Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. *Plant J.* 74:534–44
123. Tuberosa R, Turner NC, Cakir M. 2014. Two decades of InterDrought conferences: Are we bridging the genotype-to-phenotype gap? *J. Exp. Bot.* 65:6137–39
124. Tuinstra MR, Ejeta G, Goldsbrough PB. 1997. Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* 95:1005–11
125. van Dusschoten D, Metzner R, Kochs J, Postma JA, Pflugfelder D, et al. 2016. Quantitative 3D analysis of plant roots growing in soil using magnetic resonance imaging. *Plant Physiol.* 170:1176–88
126. Vilhjálmsson BJ, Nordborg M. 2013. The nature of confounding in genome-wide association studies. *Nat. Rev. Genet.* 14:1–2
127. 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:e1004845
128. Wang Q, Xie W, Xing H, Yan J, Meng X, et al. 2015. Genetic architecture of natural variation in rice chlorophyll content revealed by a genome-wide association study. *Mol. Plant* 8:946–57
129. Wang X, Wang H, Liu S, Ferjani A, Li J, et al. 2016. Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. *Nat. Genet.* 48:1233–41
130. Wang X, Wang H, Long Y, Liu L, Zhao Y, et al. 2015. Dynamic and comparative QTL analysis for plant height in different developmental stages of *Brassica napus* L. *Theor. Appl. Genet.* 128:1175–92
131. Wijnen CL, Keurentjes JJ. 2014. Genetic resources for quantitative trait analysis: novelty and efficiency in design from an Arabidopsis perspective. *Curr. Opin. Plant Biol.* 18:103–9
132. Wijnker E, Deurhof L, van de Belt J, de Snoo CB, Blankestijn H, et al. 2014. Hybrid recreation by reverse breeding in *Arabidopsis thaliana*. *Nat. Protoc.* 9:761–72
133. Wolyn DJ, Borevitz JO, Loudet O, Schwartz C, Maloof J, et al. 2004. Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana*. *Genetics* 167:907–17
134. Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLOS ONE* 4:e8451
135. Yan J, Warburton M, Crouch J. 2011. Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop Sci.* 51:433–49
136. Yang DL, Jing RL, Chang XP, Li W. 2007. Identification of quantitative trait loci and environmental interactions for accumulation and remobilization of water-soluble carbohydrates in wheat (*Triticum aestivum* L.) stems. *Genetics* 176:571–84

131. Established the idea of combining recombination suppression and genome elimination to easily generate complete sets of chromosome substitution lines.

137. Yang N, Lu Y, Yang X, Huang J, Zhou Y, et al. 2014. Genome wide association studies using a new non-parametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLOS Genet*. 10:e1004573
138. Yang Z, Huang D, Tang W, Zheng Y, Liang K, et al. 2013. Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *PLOS ONE* 8:e68433
139. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, et al. 2016. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48:927–34
140. Ye L, Huang Y, Dai F, Ning H, Li C, et al. 2015. Identification of two key genes controlling chill haze stability of beer in barley (*Hordeum vulgare* L). *BMC Genom*. 16:449
141. Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–51
142. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38:203–8
143. Yuan W, Flowers JM, Sahraie DJ, Ehrenreich IM, Purugganan MD. 2016. Extreme QTL mapping of germination speed in *Arabidopsis thaliana*. *Mol. Ecol.* 25:4177–96
144. Zhang N, Gibon Y, Wallace JG, Lepak N, Li P, et al. 2015. Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *Plant Physiol*. 168:575–83
145. Zhang W, Dai X, Wang Q, Xu S, Zhao PX. 2016. PEPIS: a pipeline for estimating epistatic effects in quantitative trait locus mapping and genome-wide association studies. *PLOS Comput. Biol.* 12:e1004925
146. Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, et al. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2:467