Rodomiro Ortiz Ríos

Plant Breeding in the Omics Era

Springer

Plant Breeding in the Omics Era

Rodomiro Ortiz Ríos

Plant Breeding in the Omics Era

Springer

Prof. Rodomiro Ortiz Ríos
Department of Plant Breeding
Swedish University of Agricultural Sciences
(SLU)
Alnarp
Sweden

ISBN 978-3-319-20531-1 ISBN 978-3-319-20532-8 (eBook)
DOI 10.1007/978-3-319-20532-8

Library of Congress Control Number: 2015949233

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

There have been significant increases in crop yields since the 1950s, which made food cheaper and more affordable. World crop production must, however, increase in the next three and a half decades to feed the ever-growing population. It should occur mostly in lands that are already under cultivation. Plant breeding, the subject of this book, provides means to address this priority global challenge.

During the twentieth century the conservation of plant genetic resources, through national, regional, and international gene banks, became well established, and as a result, major collections are available today for most crops. The use of this crop genetic endowment in plant breeding remains, however, limited due to the lack of systematic research to provide a comprehensive framework for the efficient identification and introgression of beneficial variation for both on-going priority traits and for novel added-value traits. Quantitative and population genetics are very important for germplasm conservation, genetic enhancement, and improvement of breeding methods. Understanding the types of gene action for economically important traits will improve plant breeding efficiency. Advances in omics research and in computational systems allow developing efficient approaches for plant breeding.

The first part of this book gives an overview of plant breeding and its role in producing high-yield cultivars that increase farming profitability and sustainability. Plant genetic resources and diversity are the focus on Chap. 2, which also refers to germplasm enhancement (or pre-breeding), which can be used after identifying a useful trait to "capture" its genetic diversity and put it into a "usable" form. Wild species and landrace germplasm are useful sources for developing germplasm adapted to stressful agroecosystems. Inbred lines (Chap. 3) are useful in genetic research, allele mining, or directly as cultivars in self-fertilizing species and as parents of hybrids and synthetic cultivars. Chapter 4 deals with population improvement methods such as mass and recurrent selections. Both Chaps. 3 and 4 include references about dissecting the genetics of traits or using DNA markers for introgressing or incorporating genes and quantitative trait loci.

Hybrid cultivars are among the main achievements of plant breeding in the twentieth century (Chap. 5). They ensued from exploiting heterosis, which led to a significant edible yield increase in various seed crops. Interspecific hybridization facilitated the successful introgression of wild genes into the cultigen pool. Muta-

tion breeding was used to develop cultivars of 200 species that are grown elsewhere (Chap. 6). Mutants also allow gene isolation, identification, and cloning, which can be also useful for plant breeding.

Chapter 7 provides up-to-date information on transgenic crops, which appear to perform better than their conventional counterparts in terms of yield, production costs and gross margins, and reduction in chemical pesticide use, and gives details on new breeding technology based on genetic engineering. It also argues that a regulatory system should be based on the traits of the bred crops, rather than on the method used to develop them. Genome sequencing, other omics, and synthetic biology are the topics of Chap. 8, which presents an overview on methods that reveal variation and manage them, thus assisting both crossbreeding and genetic engineering.

Examples of breeding self-fertilizing (rice, tomato, and wheat), outcrossing (cassava, cotton, and maize), and polyploid (banana/plantain and potato) crops are included in Chaps. 9, 10, and 11, respectively. These crops differ in their breeding systems, inheritance (disomic versus polysomic), propagation (sexual or vegetative), production system (annual or perennial), and use (food, feed, and fiber), whose overview provides a good conceptual underpinning of plant breeding and genetics, as well as knowledge about the sustainable use of genetic resources in crop improvement.

Chapter 12 refers to seed production, which is a key step for the success of a plant breeding program aiming cultivar development. The focus of the last chapter is on intellectual property and plant variety protection—proactively sought by those seeking rewards for innovations and believing that society welfare improves through inventions.

This book aims that the reader learns from the past and looks at the future of crop improvement. Plant breeding today, as it was before, depends on crop biodiversity and its sustainable use, which can be further facilitated by advances in omics and bioinformatics. It starts with assessing plant genetic resources (wild species, landraces, obsolete cultivars, and genetic stocks) variations aiming to enhance the cultigen pool. Research on genetics—aided nowadays by omic tools—should lead to designing knowledge-based plant breeding, which could bring further genetic gains in the breeding pools. Nonetheless, plant breeding will increasingly require pursuing a holistic interdisciplinary approach based on integrated system-oriented thinking.

Lomma, Sverige, March 2015

Rodomiro Ortiz Ríos
Faculty [Chair] Professor,
Genetics and Plant Breeding
Department of Plant Breeding
Swedish University of Agricultural Sciences

Contents

About the Author

Rodomiro Ortiz Ríos is a professor of genetics and plant breeding at the Swedish University of Agricultural Sciences (SLU) in Alnarp, Sweden. He has worked as a geneticist at Universidad Nacional Agraria La Molina (UNALM), Centro Internacional de la Papa (CIP, Lima, Perú), Rutgers University (Chatsworth, New Jersey, USA), and the International Institute of Tropical Agriculture (IITA, Ibadan, Nigeria). He held a Nordic professorship in plant genetic resources at the Danish Royal Veterinary and Agricultural University (merged now with the University of Copenhagen, Denmark), served as director at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT, Patancheru, Telangana, India), IITA, and the Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT, El Batán, México), and independent freelance consultant for international, regional, and national organizations engaged in agricultural research for development. CGIAR awarded the IITA in 1994 the prestigious King Baudouin Award for the multidisciplinary research of the team working in plantain and banana improvement, in which Ortiz was both a hands-on researcher and program leader. Volume 36 (2012) *Plant Breeding Reviews* was dedicated to him (http://media.johnwiley.com.au/product_data/excerpt/43/11183458/1118345843-61.pdf). His professional expertise includes genetics, genetic resources, germplasm enhancement, plant breeding, agro-biotechnology biosafety, intellectual property rights, and developing agroecosystem and livelihood-system approaches aiming at sustainable intensification of agriculture in the developing world. His *h*-index was 46 according to Google Scholar by mid-2015.

Chapter 1
Introduction to Plant Breeding

Plant breeding is human directed selection in genetically variable populations of plants.
William F. Tracy, Univ. of Wisconsin-Madison

Plant breeding throughout most of the twentieth century was driven by crossing parents with desired traits to generate genetic variation through recombination, and selecting the best offspring based on the phenotypes throughout generations, across locations, and over time. Since the end of the 1980s, research investments in cell and molecular biology grew significantly, whereas public plant breeding, particularly in North America and Europe, became weak (Knight 2003). In the academics, retiring professionals, who were engaged in hands-on plant breeding, were replaced by cell and molecular biologists whose main aim has been often to publish their research findings in high-impact factor journals rather than on releasing cultivars, developing segregating populations, or producing genetic stocks, which were the main tasks of their predecessors. In the last decade, decision-makers started realizing that without funding plant breeding the promise of cell and molecular biology may not contribute to developing new crop cultivars.

At present, the future of plant breeding looks more promising. Conventional crossbreeding methods are being used along with knowledge and tools ensuing from advances in omics and transgenic research. Genome sequences are also becoming available for many plant species, including the most important world's staples. As a result, more DNA markers are becoming available to facilitate modern plant breeding.

R. Ortiz Ríos, *Plant Breeding in the Omics Era,* DOI 10.1007/978-3-319-20532-8_1

Plant Breeding and Society

Plant breeding aims to provide crops that address human needs. It has been, therefore, contributing significantly to society by providing the seeds of new high-yielding cultivars with other desired traits that increase farming profitability and sustainability (Woeste et al. 2010). After the Second World War, crop yields increased steadily as a result of enhancing their genetic potential (Mifflin 2000). Modeling research shows that crop yields of most important staples feeding the world would have been 19.5–23.5% lower in 2000 without international plant breeding, which led to the Green Revolution in the developing world (Evenson and Gollin 2003). It also indicated that as a result of low yield, equilibrium prices for all crops would have been 35–66% higher in 2000 than they actually were. High-food prices could contribute to expanding crop plantings, thereby affecting the surroundings ecosystems. Without enough food supply, there would have been 13.3–14.4% lower per capita calorie intake and an increase of malnourished children between 6.1 and 7.9% in the developing world (Evenson and Gollin 2003). The Green Revolution therefore enhanced the health status of 32–42 million preschool children. In summary, most consumers worldwide benefited from the low-food prices brought by the Green Revolution as well as farmers whose crop yields rose more than food prices falls or who harvest their own food.

The Green Revolution, due to the adoption of newly bred high-yielding cereal cultivars, saved an estimated 18–27 million ha from being brought into agriculture (Stevenson et al. 2013). Likewise, although emissions from fertilizer production and use rose between 1961 and 2005, the net effect of high crop yields due to the Green Revolution avoided emissions of up to 161 Gt of carbon (GtC; 590 $GtCO_2e$) since 1961 (Burney et al. 2010). These results show that improving crop yields through plant breeding should be a must in any agricultural intensification strategy aiming to reduce greenhouse gas (GHG) emissions.

The twenty-first century has been acknowledged as the century of plant breeding, due to its expected great contribution to global food (Stamp and Visser 2012). Annual breeding gains should increase by 2.5 the current rates for doubling crop yields by 2050 (Fischer and Edmeades 2010). The environment will also benefit from plant breeding that improves crops for adapting to stressful environments or enhances host plant resistance to pathogens and pests, thus reducing negative impacts in agroecosystems such as the high use of fertilizers, pesticides, and water (Brummer et al. 2011). Plant breeding will therefore focus in this twenty-first century to develop crop cultivars that produce "more with less," that is, resource-use efficiency in a modern bio-based society. Genetic enhancement will also assist on adapting crops to the changing climate as a result of global warming and helping to mitigate GHG emissions. The newly bred crop cultivars should also contribute to conservation agriculture by improving soil health, stopping soil erosion, avoiding nutrient and chemical runoff, and enhancing biodiversity in the target agroecosystems. Plant breeding remains a cost-effective approach for improving crop productivity and to provide enough, safe, nutritious, and healthy food to the rising world population (Baenziger and Al-Otyak 2007), whose demands are increasing, especially when their wealth betters.

Sustainable intensification of agriculture (or increasing steady crop production from existing farmland) will be contributing greatly to a bio-based economy. Crops can achieve both yield increase and stability through plant breeding (Elsen et al. 2013). Yield potential and stability can be enhanced through genetic gains for this trait and by improving host plant resistance to pathogens and pests affecting crops. Adaptation to stressful environments will enhance resilience of crops. Likewise, bettering the quality and safety of food could be among achievable plant breeding aims. The genetic enhancement of crops can reduce or eliminate undesired toxic molecules or lead to cultivars showing preferred nutritional traits. Feed composition will allow both reducing grazing lands and minimizing methane gas emissions. The intrinsic traits of feedstock sources for bio-based materials may be also changed through plant breeding.

Genetics, Omics, and Plant Breeding

Plant breeding combines science, art, and business for improving crops that benefit human beings. The art of plant breeding emerged before the inheritance laws were known (Jauhar 2006). Indeed, the human ability for observation, intuition, and judgment has always determined the success of plant breeding. Likewise, managing soundly money, people, land, and time enhances the returns to investments in crop genetic enhancement. The rules of genetics made plant breeding a science-based approach for crop improvement. Plant breeding will therefore continue relying on genetic variation, selection, and evaluation of inbred lines, populations, or hybrids (Baenziger and Al-Otyak 2007). The advances in biotechnology can further accelerate the pace of plant breeding and its genetic gains. Crossing schemes based on parent's DNA "fingerprints" will provide means for improving the introduction of genetic variation, whereas the reliability of selection based on field trials may be further increased by the use of DNA markers. Advanced experimental designs and biometric methods will contribute to evaluation's accuracy and precision. Genetic engineering will provide means for introducing traits from other species that are not available in the target crop gene pool or its wild relatives, thereby supplementing novel diversity for plant breeding.

Gregor Mendel and Charles Darwin provided the foundations of today's plant breeding, that is, genetics (Mendel 1866) and hybridization and selection (Darwin 1859; Darwin and Wallace 1858), respectively. Certainly, the laws of heredity and evolution underpin plant breeding. Until the last decade of the twentieth century, plant breeding was based on incorporating information about genes (Mendelian approach) or from relatives (biometric approach) when doing selection. Irrespective of the approach used, the response to selection always indicates the genetic gain due to crossing selected parents. The genetic gain depends on the available phenotypic variation, the trait heritability (or percentage of phenotypic variation among individuals in a population attributed to their genotypes), the selection intensity (i.e., selected fraction of the population to be parents of the next generation), and the time spent for completing a selection cycle.

Mendel's laws of inheritance, which can be extended to quantitative or complex traits, provide the genetic basis of plant breeding (Arterburn et al. 2009). The main tasks on the genetic enhancement of crop are to obtain segregating populations and select for favorable allele arrangements therein. This gene reassortment should lead to plants exhibiting traits that enhance their performance, for example, high yield, host plant resistance, adaptation to stressful environments, or desired produce quality.

Since the mid-1990s, farmers have been growing transgenic crops, which ensued from pioneering work on plant genetic engineering in the early 1980s. Advances in DNA-derived technology in the last two and half decades led to marker-aided breeding (MAB). Omics research in recent years has been further contributing to identify genes and understand their functions (Moose and Mumm 2008). Today, DNA sequencing helps unraveling the relationships among alleles controlling traits. This knowledge allows establishing molecular breeding, which includes new methods and tools for assembling crop diversity, managing its genetic variation, and using it for developing new cultivars.

Populations

The Hardy–Weinberg law states that the frequency of alleles and genotypes remains constant generation after generation when evolutionary influences are absent (Hardy 1908; Weinberg 1908). Crop domestication—a human endeavor—had a significant effect on allele frequency of and the type of genetic segregation in plant populations, particularly in those loci that bear genes producing a striking morphological change. Alleles at these loci were fixed during early crop domestication, thereby reducing genetic diversity for traits controlling seed dispersal, defensive mechanisms, competing ability or plant habit, among others.

The evolution of cultivated plants could further disrupt Hardy–Weinberg equilibrium through selection, nonrandom mating, genetic drift in small populations, migration through gene flow, mutation, and meiotic drive favoring transmission of allele(s) regardless of its phenotypic expression. The domestication "bottleneck" arose when crops lost variation and changed their gene frequency by eliminating alleles at low frequency. A strong selection could, however, favor the fixation of rare alleles with a frequency below 0.05 controlling desired traits. This "bottleneck" also increased linkage disequilibrium—the nonrandom co-inheritance of alleles at different loci—because it could eliminate recombinant lineages (Hamblin et al. 2011). Patterns of linkage disequilibrium varied further among populations due to trait selection during crop evolution and breeding, which facilitates the mapping of quantitative trait loci (QTL) in cultivated plants. Divergent selection caused strong population structure or subdividing demes that may generate associations between phenotypes and unlinked markers. It can therefore provide a source for allelic diversity by using admixed populations.

Plant breeding methods continue evolving due to increasing genetic knowledge. They add in the breeding population new alleles through migration and mutation,

rearrange alleles after crossing and recombination, and remove or lose alleles through selection and random genetic drift (Cowling 2013). A sustainable approach should therefore aim to avoid losing genetic diversity in the elite breeding pool. Although migration increases genetic diversity, the use of exotic germplasm may bring negative impacts on trait performance in the elite breeding population. There are, however, germplasm enhancement (or pre-breeding) methods that may avoid such negative impacts (Ortiz 2002). These methods assist introgressing or incorporating alleles from exotic germplasm while minimizing linkage drag or bringing undesirable gene(s) that are linked to the target gene(s). Large effective population size will also assist to avoid genetic diversity loss due to selection and random genetic drift.

Analysis of genetic diversity and relationships in genebank accessions or breeding lines and populations assists germplasm enhancement (Mohammadi and Prassana 2003). The methods for accomplishing this task may rely on pedigree, divergent morphology, diverse field performance, and biochemical or DNA markers. The sampling strategy, the sort of datasets, the types of genetic distance measurements and clustering procedures, and how to establish genetic relationships are very important to ensure accurate and unbiased estimates of genetic diversity. In recent years, many user-friendly software packages, which consider evolutionary models, became available through internet and are facilitating this analysis of genetic data. These packages provide means for estimating polymorphism level, frequency of allele and genotypes, homozygosity and heterozygosity, heterogeneity or cluster patterns, fitting to expected Hardy–Weinberg ratios, and numerical resampling by either using subsets of available data (jackknifing) or drawing randomly with replacement from dataset points (bootstrapping).

Genetic Diversity

Several concepts are used to assess genetic diversity. They take into consideration the richness or the number of different forms and the evenness or equality in frequency of the different types. The allelic richness refers to the total number of distinct alleles, whereas the coefficient of gene diversity is the probability that two gametes randomly chosen from a population or sample differ at a locus.

A large genetic diversity indicates that there are a large number of alleles, especially when the variance is low, that is, evenness in frequency. Gene diversity can be calculated within a specific population and among populations of a species.

There are other genetic diversity measurements that allow knowing the structure of the variation observed. They are the heterozygosity level, Wright's fixation index F, the degree of linkage disequilibrium, and the degree of population divergence F_{ST} or G_{ST}. The heterozygosity level measures the arrangement of alleles into the genotypes while F measures the deviation of genotypic frequencies from an expected random mating or panmictic population. The linkage disequilibrium determines the arrangement of alleles at several linked loci compared with the random assortment

of gametes or zygotes, while F_{ST} or G_{ST} establish the arrangement of alleles in populations, using the variation in specific alleles among different populations. The F_{ST} measures population differentiation ensuing from population structure using biallelic DNA markers. The G_{ST} is a quantitative index of the degree of genetic differentiation between subgroups or population divergence considering multiple alleles. It ranges from zero for equal frequency of alleles to one when two populations have fixed different alleles. Total heterozygosity provides another measurement of total allelic for a species, and can be estimated by adding the allelic diversity within and among populations. The percentage of polymorphic loci has been used for measuring diversity of DNA markers. Counting the number of polymorphic or monomorphic loci and dividing this sum by the total number of loci give, in percentage, the genetic statistics. The polymorphic information content measures the ability of each DNA marker to discriminate among individuals.

The mean values of total and average heterozygosity plus the degree of population divergence for each species are obtained by summing up all polymorphic loci. The number of alleles at each polymorphic locus was defined as the total number of alleles observed divided by the number of polymorphic loci. Summing the number of alleles in each population and dividing by the total number of alleles observed within the species determines the weighted mean value of any genetic diversity measurement. These measurements within and among plant populations using isozymes or DNA markers have considered the geographic range, the mode of reproduction, the mating or breeding system, the seed dispersal mechanism, the stage of succession or life form, taxonomic status, and their domestication level. For example, plants with predominant self-fertilization (or selfing species) may show large within population genetic diversity due to the heterogeneity in the allele frequencies among their populations (i.e., some populations have low levels of genetic diversity and others show much more variation), whereas some plants with cross-fertilization (or outcrossing species) may have huge population-to-population variation.

Effective population size (Wright 1931) can be a useful measurement for both genetic resources conservation and plant breeding research because it can indicate the amount of genetic diversity of a set of individuals in a given situation (Vencovsky and Crossa 2003). It has been defined as the size of an ideal population whose genetic drift or decrease of heterozygosity rates (or increase of inbreeding) are the same as in the actual population. This measurement allows comparing diverse subpopulations, samples, offspring, or accessions with regard to an idealized reference population. For example, inbreeding slightly influences the effective population size in subpopulations, which depends mostly on the allelic diversity among them and the number of subpopulations sampled (Vencovsky and Crossa 2003). The number of seed parents and the coancestry or the degree of relationship by descent between two individuals defines the effective population size when sampling seeds in a single population with a family structure. Likewise, an effective population size kept throughout recurrent selection may significantly determine the fixation of favorable alleles. Maintaining the effective population size after selection cycles will ensure a broad genetic base that avoids reducing diversity in the breeding pool. Large populations may harbor great multigenic variation and will likely segregate

for modifier genes that can ameliorate deleterious effects brought by introducing some major gene (Walsh 2001). Codominant DNA markers can provide estimates of inbreeding, coancestry, and allelic diversity, which can be used further for calculating a reliable effective population size.

Distance Measures

The degree of similarity measured by DNA markers may allow establishing genetic relationships in plant germplasm, identifying essential derived cultivars, determining the diversity level in a gene pool, and defining heterotic groups among breeding populations and elite materials. The distances based on DNA markers are yet to prove their ability for predicting heterosis. The use of few DNA markers linked to QTL accounting for heterosis or the accumulation of favorable alleles in the hybrid are more important than the divergence among their parents per se.

There are two basic types of measurements: Euclidean and statistical distances. The Euclidean metric between two plants is a straight line measuring the "ordinary distance" as defined by the difference of the frequency of alleles between them. In a biallelic diploid population, the individual frequency differences for each allele can be 0, 0.5, or 1, which may cause the distance hard to understand since 0 may occur frequently and may mislead to the false conclusion of two plants being more similar than they really are. The "zero effect" can be dealt by considering the geometric distance calculated on a per locus basis of the total number of loci examined, which corrects the distance between the two plants. This adjusted measurement ranges between 0 (full similarity) and 1 (maximum differentiation). The other problem arises because Euclidean metric does not consider the allelic frequency variances and the relationship among the alleles. This relationship is very important because allelic frequencies are not independent within a locus or between linked loci. To solve this problem, the Euclidean distance should be corrected by dividing its value by the variance and covariance of the variable.

There are various statistical distances, whose use depends on the research aims, the properties of the DNA markers, the type of subsequent multivariate analysis, the genealogy of the germplasm, and the operational taxonomic unit, for example, clones, lines, or populations in plant breeding (Reif et al. 2005). The mathematics and genetics behind each of these measurements should be taken into account when choosing any of them for analyzing DNA marker data. The best distance measures are those that extract maximum information from DNA marker data according to the research aims, and facilitate the genetic understanding of the ensuing findings. The best DNA markers for plant population research are those from dense genetic maps such as single nucleotide polymorphisms (SNP) because they increase the precision of relatedness estimates among individuals (Weir 2007). This precision may also reveal heterogeneity along the plant genome, which may lead to having varying selfing rates across loci because of their distinct genealogy. Likewise, diversity due to genomic heterogeneity arising from mutation and selection will need to be taken into account for predicting genetic gains due to selection.

Grouping Germplasm

The best grouping strategy produces most compact and well-separated groups showing minimum variability within and maximum variability among groups (Crossa and Franco 2004). Multivariate techniques on continuous and categorical traits are used for grouping plant germplasm when more than one trait is measured in one individual or population. Univariate analysis of variance considers the variation on each trait independently, whereas multivariate procedures establish the relationships among the traits and determine how the plants (or groups of plants) vary when considering all traits together.

The nonhierarchical principal component analysis (PCA) determine common patterns of variation among groups and subgroups of genebank accessions or breeding materials based on the variance/covariance structure, with a few (usually two or three) linear combinations of the original variables. PCA reduces the dimension of a dataset between two individuals from 2 × trait number to 2 × 2 (when considering only two principal components) or 2 × 3 (for three principal components). Hence, two or three values replace all traits since they are regarded to capture the differences between the two individuals. Principal components (or PRIN = Principal component coefficients) are functions of the eigenvalues and eigenvectors of the variance/covariance matrix. They are often used in the development of a discriminant function to assign clones to each taxonomic group. PCA needs variables that follow a multivariate normal distribution, which is not the case with DNA marker data. If PCA uses raw marker data such as the popular scoring of presence or absence of bands after gel electrophoresis or DNA fragment analysis, then it will have a distorted matrix because variance and covariance are based only in two or three values, and its PRINs are not regarded as independent. PCA must be therefore run on genetic distances that are calculated from DNA marker data.

Cluster analysis is a hierarchical procedure—also known as compact linkage analysis but without inferring a genetic test—that groups genebank accessions or breeding materials individually. Clusters are merged sequentially based on distance measures using an algorithm that initially uses each genebank accession or breeding material as a cluster. The diagrammatic depictions of eigenvalues are shown on a dendrogram, which is a tree-like diagram placing individuals with close-distance measurements nearby because they share similar phenotypes or DNA marker "fingerprints."

Quantitative Variation

The phenotypic variation of a plant population measured across locations, seasons, or years can be attributed to its genetics, the environment where it grows, and the genotype–by–environment interaction (GE). Biometric models are used to explain traits with continuous variation because algebraic equations facilitate the

understanding of their quantitative genetics, which is the study of complex traits affected by the action of multigenes. Quantitative genetic models include the various genes—having major or small effects—and the nongenetic factors affecting a complex trait. These models were influenced by the early research of Sir Ronald A. Fisher (1918) and Sewall G. Wright (1921) on the analysis of variance components and the resemblance between relatives, respectively, as well as by the mathematical theory of natural and artificial selection of J.B.S. "Jack" Haldane (Haldane 1932). Quantitative genetics along with biometrics were used extensively during the previous century to determine the number of loci controlling quantitative trait variation, the nature of QTL-bearing alleles with a range of effects, the types of gene action and their additive and nonadditive effects, epistasis, and GE (Lamkey and Lee 1993), which are very relevant when breeding complex traits in crops, for example, edible yield.

Maize has a long history as a model genetic system since the early decades of the last century (Wallace et al. 2014). Plant breeding research based on the quantitative genetics has shown the preponderance of additive genetic variance in maize, which indicated that genetic gains were primarily due to selection of favorable alleles with additive genetic effects (Hallauer 1980). This genetic knowledge led to apply recurrent selection aiming to increase the frequency of favorable alleles in breeding populations of this outcrossing species. By using this cyclical selection program, offspring are evaluated in replicated trials, and those with superior breeding values are recombined to obtain offspring for further selection. This approach allows to simultaneously improving the mean performance and maintaining genetic variation in the improved population, which can be also used as a source for extracting inbred lines for hybrid breeding.

The efficiency of selection for inbred lines from selfing species, especially when using pedigree-based breeding methods, can be improved by a quantitative genetics approach. The best linear unbiased predictor (BLUP), which was initially devised for animal breeding, includes information regarding relationships among the offspring and estimates random genetic effects of a mixed model (Piepho et al. 2008). This method may provide, with great accuracy, estimates of breeding values (or the genetic merit of an individual based on its ability to produce superior offspring), thus likely enlarging genetic gains from selection (Ramalho et al. 2013). BLUP can be also used for predicting hybrid performance of outcrossing species, and for modeling the GE.

The estimates of gene action and their effects according to the most quantitative genetic models were based on averages over the whole genome rather than on individual loci. These models seldom included the interaction between nonallelic genes or epistasis because it was difficult to estimate or mathematically intractable (Walsh 2001). A genotype may not be a very accurate predictor of phenotype when this interaction and the GE are significant.

The underlying basis of a phenotype is known as genetic architecture. Quantitative trait variation may display a complex genetic architecture because there are few genes with large effects or many genes with small effects involved, and these genes can show additive, dominance, or epistatic effects, and interact with

the environment. Hence, each gene effect magnitude may vary significantly. These genes, which may be distributed across the genome randomly or following a certain pattern, can also affect quantitative trait variation through pleiotropy, that is, single genetic variation affects various phenotypic traits.

Mapping Traits

Genetic mapping depends on trait heritability, sample size, and the genetic dissimilarity among individuals included in the sample. Although the idea of mapping QTL in plants dates back to the 1920s (Sax 1923), the availability of genetic markers was its main limiting step. It was in the 1980s that QTL linkage analysis began in crops to determine with some reliability level what markers in a chromosome or genome region were able to account for the dissimilarity of phenotypes among genetically related individuals of the mapping population (Tanksley et al. 1982). Cosegregation of the variants of the marker and the alleles at the trait locus allowed this linkage analysis. Analyses based on dense DNA marker maps, which were mostly constructed using microsatellites (SSR) and SNP, facilitated further the understanding of the genetic architecture of these traits and identifying genes with large effects or many dozens of genes that account for some of the phenotypic variation (Hill 2010). Likewise, recent progress made in plant genomics—and more recently in DNA sequencing of many crops—coupled with the availability of biometric methods for analyzing genetic and phenotypic data with friendly software, made feasible to map and dissect complex quantitative trait variation (Posthuma et al. 2003). These analyses often found many loci contributing individually a small amount to this variation.

Linkage disequilibrium using historical recombination events provides another means for identifying associations between variation of target traits and polymorphic DNA markers (Hill 2012). This approach does not need making any specific mating for developing experimental mapping populations, which are often time consuming and expensive. Association analysis can use data from available nursery or advanced breeding trials, and multi-environment testing. The distance between loci across chromosomes defines the linkage disequilibrium, which has been very useful for dissecting complex quantitative trait variation based on fine-scale mapping and historical recombination. False positive associations between DNA markers and target traits may, however, occur when significance tests lack stringency, or due to population structure ensuing from admixture, mating system, genetic drift, selection, and a low frequency of alleles in the initial population. Separating linkage disequilibrium due to physical linkage from that arising from population structure is therefore a must before doing association analysis. Bayesian analysis, and clustering or scaling account for population structure, being the former the most effective for assigning individuals to subpopulations using unlinked DNA markers. The evidence about a true state is given in terms of degrees of belief in Bayesian statistics, which evaluates the probability of a hypothesis based on a prior probability that updates according to new, relevant data.

Appropriate biometric methods assist identifying polymorphisms that affect quantitative trait variation in a population. A sound association analysis includes the relatedness between individuals in the population and the modeling of the GE (Crossa et al. 2007). The coefficient of parentage, the DNA marker-based estimation of the probability of identity by descent between individuals, or both simultaneously can detect the degree of relatedness. Linear mixed models based on phenotypic data allow the accurate prediction of genotypic performance when using covariance structures defined by the genetic association between relatives participating in the experiment. Any suitable biometrical model assessing the DNA marker–trait association should always consider the population structure and covariance among relatives, plus the interactions of the environments with DNA markers, subpopulations, and lines or clones nested within subpopulations, as well as their respective main effects.

Quantitative genetics can now undertake comprehensive large-scale analyses due to the availability of high-throughput omics methods (Keurentjes et al. 2008). This shift from dealing with single traits enables to research on how genetic information translates into biological function(s), including both transcriptional and (post)translational regulation, plus metabolic signaling pathways. This kind of research facilitates unraveling regulatory networks that integrate biological information flow in the gene-to-function pathway. The joint use of linkage disequilibrium mapping and transcriptomics may provide means for identifying regulatory genetic factors affecting quantitative trait variation in plants. The ensuing knowledge will assist an understanding of how QTL operates and how it is regulated.

Genotype–by–Environment Interaction

Breeding materials should be included in multi-environment testing for a thorough appraisal of their performance (or phenotype). The phenotypic effects of the interactions between genotypes and the environments where they grow are known as the GE. GE may be noted as a change in ranking of genotypes across environments or in the relative magnitude of their gene effects in response to the environments. Understanding the contributions of the GE to crop performance will provide knowledge for its appropriate design of testing and selection. Hence, GE needs to be appropriately determined using sound biometric methods to guide decisions in plant breeding.

Either the genotype or the environment can be fixed but the other should be regarded as random when studying GE using a linear model. All levels of the populations of parameters are included for the fixed effect while the random factor only takes a random sample of population levels (Basford et al. 2004). The genotypes are usually regarded to be a random sample from the breeding population, and the managed testing environments are often fixed since repeated across years and locations, thus defining a mixed model. There are various models for gaining insights and predicting the GE (Malosetti et al. 2013). The descriptive models group genotypes and environments, whereas other models explain GE using covariates that assist the modeling.

Factorial and partial least squares regressions incorporate external environmental and genotypic covariables directly into the model for interpreting GE (Vargas et al. 1999). Factorial regression is an ordinary linear model that allows the inclusion of external variables such as crop husbandry, soil, or weather data. These variables could, however, show a high collinearity— that is, the predictor variables of a multiple regression model are highly correlated—thus complicating the interpretation of the least square regression coefficients. The partial least regressions, which are bilinear models, offer a solution to this multicollinearity and describe GE according to the differential sensitivity of genotypes to the environmental covariables that are linear combinations of the complete set of measured environmental variables.

Modeling helps to visualize complex data and increases accuracy. The additive main effects and multiplicative interaction (AMMI) model is used for two-way data tables ensuing for multi-environment trials (Gauch 2006). The AMMI model accounts first for the main effects and then uses the PCA for analyzing the interactions. Exploratory scatter or dispersion graphs such as the bi-plots are used widely for assessing GE. For example, the AMMI1 bi-plot shows the main effects for the genotypes in the horizontal axis, whereas the main effects for the environments are along the vertical axis. The GE for a given genotype and environment is estimated by multiplying their respective scores. Positive GE occurs when both have the same sign for these scores, but it is negative if they have opposite signs. The genotype main effects and genotype–by–environment interaction effects (GGE) model also visualizes patterns in trial (Yan et al. 2000). GGE graphs show what genotype performs best where, and efficiently assess the representativeness and discriminating ability of testing locations, thus defining mega-environments. Mega-environments are broad, often discontinuous, environments having same biotic and abiotic stresses, cropping systems, consumer preferences, and levels of production. The GGE bi-plot can also assist on visualizing the representativeness and repeatability of testing environments (Yan et al. 2011), which should discriminate genetic differences among genotypes included in the trial and represent the target environments where selected genotypes can repeat their performance.

Multi-environment trials often include as testing genotypes related breeding lines and cultivars, for example, full- or half-sibs. Data taken from these related genotypes are therefore correlated. Mixed linear models allow dealing with heterogeneous and correlated variance–covariance structures. Coefficients of coancestry and covariance matrix of breeding values should be taken into account by the linear model analyzing multi-environment trial data (Crossa et al. 2006). This variance–covariance matrix of breeding values in selfing species can be divided into additive effects, additive × additive effects and their interaction with the environment (Burgueño et al. 2007). This approach will enhance genetic gains because it helps in identifying breeding lines with high-additive effects that will be further used for crossing. The related genotypes may be also visualized when using bi-plots.

Research about the genetic basis of the GE should consider the modeling of QTL expression as influenced by the environment, which will provide important knowledge for DNA MAB schemes. A QTL lacking GE can be used across environments (wide adaptation), whereas a QTL with a significant GE can be only used in the environment where it was detected (specific adaptation). The QTL ×

environment interaction appears to be ubiquitous in most crops. Factorial regression models can be generalized to account for QTL expression dependent on environmental covariables (Malosetti et al. 2004). QTL modeling translates DNA marker information into genetic predictors. Tests for their regression coefficients assess the effects of QTL expression and the QTL × environment interaction (Malosetti et al. 2013). The ensuing QTL models can assist predicting GE arising when testing new genotypes in other environments. Mixed linear models are also suitable for analyzing GE when using association mapping (Saïdou et al. 2014). Large sample size should be used to ensure a rigorous model selection and powerful appraisal for the interactions. Likewise, Bayesian statistics and genome-wide marker information are useful for testing and estimating QTL main effects and the QTL × environment interactions (Zhao and Xu 2012). Large genome coverage by DNA markers avoids missing QTL.

Genomics research shows that GE often results by changes in the magnitude of the gene effects in response to the environment and may be associated with various genetic factors and molecular variants (Des Marais et al. 2013). Genes do not show equal GE because of their differential regulation by the environment. Whole-genome approaches can simultaneously monitor the effects of a polymorphism on thousands of genomics loci, and reveal the underlying principles, mechanisms, and evolutionary impacts of GE. It seems that GE results from changes in upstream regulators rather than local changes to promoters (Grishkevich and Yanai 2013). Generalized genetical genomics can further identify GE in trait metabolism at the molecular level (Joosen et al. 2013). This research strategy requires studying biological systems across various environments, and combines genetic and sensibly chosen environmental perturbations to understand the plasticity of molecular networks.

Phenotyping

Phenotyping remains as a significant bottleneck that limits the power of genetic analysis and genomic prediction in plant breeding. Appropriate phenotypic assessment facilitates dissecting data according to both genotypic and environmental variables, and will assist plant breeding for developing new cultivars.

In the last century, Sir Ronald A. Fisher (1925a, 1925b) contributed significantly to the theory of experimental design and statistical estimation with the aim of making meaningful tests for comparing quantitative measurements. He also emphasized the importance of both randomization and replication to increase accuracy and precision in data recording (Fisher 1935), which are still key for any phenotyping strategy aiming to empower high-resolution mapping, association genetics, and genomic selection (Cobb et al. 2013). Next-generation phenotyping aims, therefore, to enhance the accuracy, precision, and throughput of any phenotypic assessment simultaneously reducing costs and minimizing work using automation, remote sensing, improved data integration, and sound experimental design, which requires interdisciplinary undertakings involving biology, information technology, bioinformatics, biometry, and engineering.

Expensive and time-consuming phenotyping imposes limits on the sample size. Hence, selective phenotyping has been advocated to select individuals that maximize genotypic dissimilarity (Jin et al. 2004). Prior knowledge on the genetic architecture of target trait(s) makes this phenotyping approach more effective because it helps in focusing on specific genetic regions. Selective phenotyping may include genotypes that maximize the overall mapping information content in the selected offspring, or those that maximize it as well as its uniformity across the genome (Jannink 2005). Either strategy decreases error and increases the accuracy of QTL mapping even for QTL with small effect or when using DNA markers with spacing below 10 cm.

Phenomics

Precision phenotyping is a technology, still under development, that provides means to accelerate the understanding of genes and their environmental responses (Arvidsson et al. 2011; Furbank and Tester 2011). Single plant phenotyping based on robotics and image analysis offers the opportunity for precise plant development research to relate the phenotype with the genotype in controlled or semi-controlled environments.

The phenome refers to the expression of the species genome in a given environment (Furbank and Tester 2011). Plant phenomics is the study of the phenome and how it is determined over time. It closes the gene–genotype loop by facilitating trait and gene identification as well as providing insights into the genotype development process. Forward phenomics uses high throughput and fully automated and low resolution, followed by higher-resolution, lower-throughput measurements to screen germplasm for valuable physiological traits. Reverse phenomics dissects in detail valuable traits that reveal mechanistic understanding and allow exploitation of this mechanism by plant breeding. There are various automatic high-throughput plant growth and phenotyping platforms available (Dwivedi et al. 2013). These assays may be refined further to speed up comprehending gene functions and environmental responses.

High-throughput, rapid, and cost-effective phenomic platforms are still lacking for measuring accurately in the field of plant growth and development, and assess response to stress on large sets of individuals. The most powerful of them are empirical rather than analytical, and depend on large data acquisition and further processing (Cabrera-Bosquet et al. 2012). Field-based high-throughput phenotyping should also deal with the inherent spatial and temporal variability of field trials. Digital phenotyping of plant development in the field lacks an efficient imaging equipment and space to evaluate many accessions under various treatments (White et al. 2012). Precision agriculture tools can be adapted for their early use in the growing season or in small plots (Montes et al. 2007), thus enabling rapid and semiautomated measurement of traits that are of particular relevance to plant breeding.

References

Arterburn MK, Jones SS, Kidwell KK (2009) Plant breeding and genetics. In encyclopedia of life support systems. Soils, plant growth and crop production, vol 1. EOLSS Publishers, Oxford. http://www.eolss.net/sample-chapters/c10/e1-05a-08-00.pdf

Arvidsson S, Pérez-Rodríguez P, Mueller-Roeber B (2011) A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. New Phytol 191:895–907

Baenziger SP, Al-Otyak SM (2007) Plant breeding in the 21st century. Afr Crop Sci Conf Proc 8:1–3

Basford KE, Federer WT, DeLacy IH (2004) Mixed model formulations for multi-environment trials. Agron J 96:143–147

Brummer CE, Barber WT, Collier SM, Cox TS, Johnson R, Murray SC, Olsen RT, Pratt RC, Thro AM (2011) Front Ecol Environ 9:561–568

Burgueño J, Crossa J, J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A (2007) Modeling additive × environment and additive × additive × environment using genetic covariances of relatives of wheat genotypes. Crop Sci 47:311–320

Burney JA, Davis SJ, Lobell DB (2010) Greenhouse gas mitigation by agricultural intensification. Proc Natl Acad Sci U S A 107:12052–12057

Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Araus JL (2012) High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J Integr Plant Biol 54:312–320

Cobb NJ, DeClerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126:867–887

Cowling (2013) Sustainable plant breeding. Plant Breed 132:1–9

Crossa J, Franco J (2004) Statistical methods for classifying genotypes. Euphytica 137:19–37

Crossa J, Burgueño J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A (2006) Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci 46:1722–1733

Crossa J, Burgueño J, Dreisigacker S, Vargas M, Herrera S, Lillemo M, Singh RP, Trethowan R, Franco J, Warburton M, Reynolds M, Crouch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889–1913

Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London

Darwin C, Wallace AR (1858) On the tendency of species to form varieties, and on the perpetuation of varieties and species by natural means of selection. J Proc Linn Soc Lond Zool 3:46–50

Des Marais DL, Hernandez KM, Juenger TE (2013) Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. Annu Rev Ecol Evol Syst 44:5–29

Dwivedi SL, Sahrawat KL, Upadhyaya HD, Ortiz R (2013) Food, nutrition and agrobiodiversity under global climate change. Adv Agron 120:1–128

van Elsen A, Ayerdi Gotor A, de Vicente C, Traon D, Gennatas J, Amat L, Negri V, Chable V (2013) Plant breeding for an EU bio-based economy. JRC scientific and policy reports, joint research centre, institute for prospective technological Studies. European Commission, Publications Office of the European Union, Luxembourg

Evenson RE, Gollin D (2003) Assessing the impact of the green revolution, 1960 to 2000. Science 300:758–761

Fischer RA, Edmeades GO (2010) Breeding and cereal yield progress. Crop Sci 50:85–98

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb 52:399–433

Fisher RA (1925a) Statistical methods for research workers. Oliver & Boyd, Edinburgh