Trends in Genetics

Review

Broadening Our Portfolio in the Genetic Improvement of Maize Chemical Composition

Weiwei Wen,[1,2] Yariv Brotman,[1,3] Lothar Willmitzer,[1]
Jianbing Yan,[2] and Alisdair R. Fernie[1,*]

The adoption of recombinant inbred line and introgression line populations, as well as the study of association mapping panels, has greatly accelerated our ability to identify the genes underlying plant phenotypic variance. In tandem, the development of metabolomics approaches has greatly enhanced our ability to comprehensively define cellular chemical composition. As a consequence, breeding for chemical composition is being extended beyond our traditional targets of oil and protein to include components such as essential amino acids, vitamins, and antioxidant secondary metabolites with considerable purported consequences for human health. Here, we review the above-mentioned developments paying particular attention to the genetic architecture of metabolic traits as well as updating the perspective for utilizing metabolomics in maize improvement.

Trends

Extensive breeding resources are currently available for maize.

Concurrently, our capacity for evaluating chemical composition of plants has greatly increased given the advent of metabolomics.

Taken together, these advances allow us to extend our portfolio for compositional quality oriented breeding.

The Long History of Plant Breeding

Plant breeding encompasses the creation, selection, and fixation of superior plant phenotypes in the development of improved cultivars suited to the needs of humans and has been a human pursuit since the advent of cultivation. The primary goals of crop plant breeding has focused on improved yield with considerable success being obtained by the development of hybrid maize (*Zea mays L.* ssp. *Mays*; [1]), introduction of the wheat (*Triticum aestivum*) and rice (*Oryza sativa*) varieties that enabled the Green Revolution [2], as well as the molecular marker assisted **introgression** (see Glossary) of defined genes or genomic regions from wild species and landraces [3]. However, the roots of plant breeding vastly predate these modern approaches. For example, prehistoric selection for visible phenotypes, which facilitated harvesting and increasing productivity, led to the domestication of the first crop varieties [4]. Additionally, revolutionary insights made both by Darwin and Mendel over 100 years ago paved the way towards the scientific approach to plant breeding [5]. However, these insights were only adopted in earnest once a better understanding of quantitative genetics was able to reconcile Mendelian principals with continuous trait variation [5,6]. Since this was achieved, successive iterations have adopted molecular biology, modern breeding technology such as **marker-assisted selection** (MAS), and most recently genomics [7–11] to further broaden the scientific basis of plant breeding. Here, we will focus on the application of these approaches to improving the chemical composition of maize focusing in parallel on the recent advances in crop genetics and methods for chemical analysis, and their application to better understand tolerance and resistance mechanisms, as well as improving human nutrition.

Maize Genetic Diversity

Maize is not only of global importance as a food and source of diverse industrially important products but is also a model system with tremendous genetic diversity. Maize was domesticated

[1] Max Planck Institute of Molecular Plant Physiol., 14476 Potsdam Golm, Germany
[2] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China
[3] Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel

*Correspondence:
fernie@mpimp-golm.mpg.de
(A.R. Fernie).

http://dx.doi.org/10.1016/j.tig.2016.05.003

from the wild progenitor teosinte (*Zea mays* ssp. *Parviglumis*) approximately 10 000 years ago in Mexico [12]. Subsequent to domestication, maize landraces had been subject to intensive improvement efforts, culminating in the development of hybrid maize lines that are highly adapted to modern agricultural practices [13]. Teosinte is extremely diverse and maize retains much of the diversity of its wild ancestor, with any two maize varieties differing from one another by 1.4% at the DNA level [14]. It was documented that the level of nucleotide diversity found in maize is two to five times higher than that of other domesticated grass crops and is 14 times higher than that of humans [15]. Within the diverse germplasm, maize inbred lines represent a fundamental resource for studies in genetics and breeding.

The genome sequencing of a large number of maize lines and wild relative lines in recent years has provided significant insights to our understanding of the maize genome and evolution [16–18]. The B73 reference genome is 2.3 Gb in total size and is composed of approximately 85% repetitive sequences. In addition to single nucleotide polymorphisms (SNPs) and small insertion and deletions (indels), there are a large number of structural variants in the maize genomes, including causal polymorphisms for phenotypic variations. These advances in maize genomics further revolutionized our understanding of genetic diversity and provided an important foundation for designing strategies for maize improvement.

Most agriculturally and economically important traits have complex genetic underpinnings (i.e., determined by multiple **quantitative trait loci**, **QTLs**). Precisely locating and characterizing these functional loci facilitates crop improvement via MAS or biotechnology aided breeding. To dissect complex traits, **linkage analysis** and **association mapping** are commonly used. On the basis of the diverse maize germplasm, a variety of populations for the above-mentioned genetic analysis have been created or assembled. Here, we provide a short overview of different types of mapping populations with a discussion of their utilizations in maize in the following sections. For linkage analysis, recombinant inbred line (RIL) populations derived from two parental maize lines are frequently used. A noteworthy biparental population for linkage mapping is the **intermated B73 × Mo17 (IBM) population**. The Maize Mapping Project (MMP) constructed a genetic map (IBM2) that contains 2026 markers. The genetic map, the open access to seeds, and the available web resources have led to the wide use of IBM by the maize genetics community [19]. The adoption of linkage analysis provided great power and allowed the identification of epistatic interacting loci and loci exhibiting only minor effects [20–22]. However, this approach exhibits a relative paucity of alleles and unless huge mapping populations are used the mapping resolution is generally low. By contrast, association mapping possesses advantages over linkage analysis in terms of mapping resolution: allele richness and time investment. However, the power to detect minor effect loci and epistatic interactions using association mapping is often limited due to the complex genetic background of the population. Hence, the **germplasm collection** that encompasses sufficient genetic diversity covering most variations for the traits of interest and with rapid **linkage disequilibrium** (**LD**) decay is vital for a promising outcome of association mapping [23]. Linkage and association analysis can be complementary to each other and in some cases both have been used for cross-validation and causal genetic variant identification [22,24]. For the sake of combining the advantages and eliminating the disadvantages of linkage and association analysis, the **nested association mapping** (**NAM**) population that combines 25 RIL populations with 200 lines per family [25] and the **multiple parent advanced generation inter-cross** (**MAGIC**) population were developed [26]. These multiple biparental family sets and multiple parent populations were constructed to enable higher mapping resolution and power by introducing more recombination events and eliminating the confounding effects of population structure [25,26]. Compared with NAM, the MAGIC design does not use a common reference parent for all of the crosses and avoids the confounding effect of family structure on QTLs inheritance, which makes it more statistically efficient and easier to detect QTLs that contribute to differences among biparental families.

Glossary

Association mapping: also known as linkage disequilibrium mapping or association analysis; is a method that identifies the link between genomic variants and phenotypes, which takes advantage of historic linkage disequilibrium to detect and locate QTLs.

Genome-wide association study (GWAS): is an examination of associations between common genetic variants (usually genome-wide SNPs) and traits such as human diseases.

Genomic prediction: also known as genomic selection where genetic markers covering the whole genome of the training and breeding populations and the phenotypic data of the training population are used and integrated in a model to predict the performance of the breeding population. The selection decision will be made on the breeding population based on the breeding values.

Germplasm collection: collection of living genetic resources such as seeds or tissue that are maintained for the purpose of animal and plant breeding, preservation, and other research uses.

Illinois Long-Term Selection Experiment: an experiment initiated and conducted by researchers at the University of Illinois from 1896 for selecting maize lines with the highest or lowest concentrations of grain protein or oil. This study has created 12 populations that vary significantly in their grain protein and oil composition through 110 cycles of recurrent selection over a century.

Intermated B73 × Mo17 (IBM) population: the two inbred lines B73 and Mo17 were crossed to make the F1 hybrid and was then self-pollinated. F2 progenies were then intermated for four generations, followed by repeated selfing to generate recombinant inbred lines.

Introgression: the incorporation of genes from one species into the gene pool of a second species by hybridization and backcrossing. Introgression lines of a certain crop contain a genetic component artificially derived from a wild relative population through repeated backcrossing, which are used for gene or QTL mapping and breeding new varieties.

Linkage analysis: a tool for genetic mapping where the coinheritance of