Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Jeffrey Bennetzen · Sherry Flint-Garcia
Candice Hirsch · Roberto Tuberosa *Editors*

The Maize Genome

Springer

Compendium of Plant Genomes

Series editor

Chittaranjan Kole, Raja Ramanna Fellow, Department of Atomic Energy, Government of India, ICAR-National Research Center on Plant Biotechnology, New Delhi, India

Whole-genome sequencing is at the cutting edge of life sciences in the new millennium. Since the first genome sequencing of the model plant Arabidopsis thaliana in 2000, whole genomes of about 70 plant species have been sequenced and genome sequences of several other plants are in the pipeline. Research publications on these genome initiatives are scattered on dedicated web sites and in journals with all too brief descriptions. The individual volumes elucidate the background history of the national and international genome initiatives; public and private partners involved; strategies and genomic resources and tools utilized; enumeration on the sequences and their assembly; repetitive sequences; gene annotation and genome duplication. In addition, synteny with other sequences, comparison of gene families and most importantly potential of the genome sequence information for gene pool characterization and genetic improvement of crop plants are described.

Interested in editing a volume on a crop or model plant? Please contact Dr. Kole, Series Editor, at ckole2012@gmail.com

More information about this series at http://www.springer.com/series/11805

Jeffrey Bennetzen · Sherry Flint-Garcia
Candice Hirsch · Roberto Tuberosa
Editors

The Maize Genome

Springer

Editors
Jeffrey Bennetzen
Department of Genetics
University of Georgia
Athens, GA, USA

Sherry Flint-Garcia
USDA-ARS
Columbia, MO, USA

Candice Hirsch
Department of Agronomy
and Plant Genetics
University of Minnesota
St. Paul, MN, USA

Roberto Tuberosa
Department of Agricultural
and Food Sciences
University of Bologna
Bologna, Italy

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-319-97426-2 ISBN 978-3-319-97427-9 (eBook)
https://doi.org/10.1007/978-3-319-97427-9

Library of Congress Control Number: 2018950816

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book series is dedicated to
my wife Phullara, and our children
Sourav, and Devleena
Chittaranjan Kole

Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of 'markers' physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, the absence of epistasis, and codominant nature. An array of other hybridization-based markers, PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits, and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period, a number of new mapping populations beyond F2 were utilized and a number of computer programs were developed for map construction, mapping of genes, and mapping of polygenic clusters or QTLs. Molecular markers were also used in studies of evolution and phylogenetic relationship, genetic diversity, DNA-fingerprinting, and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still, they remained 'indirect' approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

Physical mapping of genomes was the obvious consequence that facilitated development of the 'genomic resources' including BAC and YAC libraries to develop physical maps in some plant genomes. Subsequently, integrated genetic–physical maps were also developed in many plants. This led to the concept of structural genomics. Later on, emphasis was laid on EST and transcriptome analysis to decipher the function of the active gene sequences leading to another concept defined as functional genomics. The advent of techniques of bacteriophage gene and DNA sequencing in the 1970s was extended to facilitate sequencing of these genomic resources in the last decade of the twentieth century.

As expected, sequencing of chromosomal regions would have led to too much data to store, characterize, and utilize with the-then available computer software could handle. But development of information technology made the life of biologists easier by leading to a swift and sweet marriage of biology and informatics, and a new subject was born—bioinformatics.

Thus, evolution of the concepts, strategies, and tools of sequencing and bioinformatics reinforced the subject of genomics—structural and functional. Today, genome sequencing has traveled much beyond biology and involves biophysics, biochemistry, and bioinformatics!

Thanks to the efforts of both public and private agencies, genome sequencing strategies are evolving very fast, leading to cheaper, quicker, and automated techniques right from clone-by-clone and whole-genome shotgun approaches to a succession of second generation sequencing methods. Development of software of different generations facilitated this genome sequencing. At the same time, newer concepts and strategies were emerging to handle sequencing of the complex genomes, particularly the polyploids.

It became a reality to chemically—and so directly—define plant genomes, popularly called whole-genome sequencing or simply genome sequencing.

The history of plant genome sequencing will always cite the sequencing of the genome of the model plant Arabidopsis thaliana in 2000 that was followed by sequencing the genome of the crop and model plant rice in 2002. Since then, the number of sequenced genomes of higher plants has been increasing exponentially, mainly due to the development of cheaper and quicker genomic techniques and, most importantly, development of collaborative platforms such as national and international consortia involving partners from public and/or private agencies.

As I write this preface for the first volume of the new series 'Compendium of Plant Genomes,' a net search tells me that complete or nearly complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and three basal plants is accomplished, the majority of which are in the public domain. This means that we nowadays know many of our model and crop plants chemically, i.e., directly, and we may depict them and utilize them precisely better than ever. Genome sequencing has covered all groups of crop plants. Hence, information on the precise depiction of plant genomes and the scope of their utilization is growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated Web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful both to students and to teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is of interest not only for the geneticists and breeders but also for practitioners of an array of plant science disciplines, such as taxonomy, evolution, cytology,

physiology, pathology, entomology, nematology, crop production, biochemistry, and obviously bioinformatics. It must be mentioned that information regarding each plant genome is ever-growing. The contents of the volumes of this compendium are therefore focusing on the basic aspects of the genomes and their utility. They include information on the academic and/or economic importance of the plants, description of their genomes from a molecular genetic and cytogenetic point of view, and the genomic resources developed. Detailed deliberations focus on the background history of the national and international genome initiatives, public and private partners involved, strategies and genomic resources and tools utilized, enumeration on the sequences and their assembly, repetitive sequences, gene annotation, and genome duplication. In addition, synteny with other sequences, comparison of gene families, and, most importantly, potential of the genome sequence information for gene pool characterization through genotyping by sequencing (GBS) and genetic improvement of crop plants have been described. As expected, there is a lot of variation of these topics in the volumes based on the information available on the crop, model, or reference plants.

I must confess that as the series editor, it has been a daunting task for me to work on such a huge and broad knowledge base that spans so many diverse plant species. However, pioneering scientists with lifetime experience and expertise on the particular crops did excellent jobs editing the respective volumes. I myself have been a small science worker on plant genomes since the mid-1980s and that provided me the opportunity to personally know several stalwarts of plant genomics from all over the globe. Most, if not all, of the volume editors are my longtime friends and colleagues. It has been highly comfortable and enriching for me to work with them on this book series. To be honest, while working on this series I have been and will remain a student first, a science worker second, and a series editor last. And I must express my gratitude to the volume editors and the chapter authors for providing me the opportunity to work with them on this compendium.

I also wish to mention here my thanks and gratitude to the Springer staff, Dr. Christina Eckey and Dr. Jutta Lindenborn in particular, for all their constant and cordial support right from the inception of the idea.

I always had to set aside additional hours to edit books besides my professional and personal commitments—hours I could and should have given to my wife, Phullara, and our kids, Sourav, and Devleena. I must mention that they not only allowed me the freedom to take away those hours from them but also offered their support in the editing job itself. I am really not sure whether my dedication of this compendium to them will suffice to do justice to their sacrifices for the interest of science and the science community.

Kalyani, India Chittaranjan Kole

Preface to "The Maize Genome" Volume

It is now three decades since the mapping of QTLs for agronomic traits, including yield, was first reported in maize. Following this pioneering and groundbreaking work, the pace of progress in maize genomics and its breeding applications have been nothing short of spectacular. This progress continued to accelerate, as witnessed by the publication of the first assembly of the maize genome a decade ago. This second milestone paper prompted and paved the way to a wealth of manuscripts and the discovery of several genes/QTLs with a relevant role in maize growth and field performance.

Based upon this premise, this volume builds on such knowledge and provides a glimpse into some of the recent advances in the study and characterization of maize genome structure, evolution and function, and how this information can be harnessed to enhance the effectiveness of genomics-assisted breeding as well as gene/QTL cloning and study. Suitable platforms, genetic materials, and databases now bridge forward and reverse genetics approaches and allow for an unprecedented level of genetic and functional resolution, particularly for quantitative traits. Maize genomics now provides breeders with a formidable toolbox for tailoring hybrids better adapted to face the challenges posed by climate change, while ensuring an environmentally sustainable and profitable production of one of the most important crops for mankind.

Overall, the chapters in this volume emphasize the importance of deeply characterizing the maize genome in order to identify rare haplotypes with beneficial effects that are not yet represented in elite germplasm. Large-scale resequencing coupled with an equally deep analysis of the transcriptome, proteome, and metabolome will accelerate the cloning of agronomically valuable loci, paving the way to a more effective harnessing of biodiversity, more accurate modeling, and, most importantly, the fine-tuning of key sequences via gene editing.

We hope that this volume will provide maize scientists with a better appreciation of the complexity underpinning phenotypic variability while stimulating their curiosity and interest in undertaking new studies to further enhance our understanding of such complexity.

The editors are grateful to the authors of the different chapters for reviewing the published research work in their area of expertise and, in some cases, sharing their unpublished results to update the articles. We also

appreciate their cooperation in meeting the deadlines and in revising their manuscripts, whenever required. This notwithstanding, the editors remain responsible for any errors that inadvertently might have crept in during the editorial work.

Athens, USA — Jeffrey Bennetzen
Columbia, USA — Sherry Flint-Garcia
St. Paul, USA — Candice Hirsch
Bologna, Italy — Roberto Tuberosa

Contents

Part I

Genome Sequencing and Genotyping

1 Draft Assembly of the F2 European Maize Genome Sequence and Its Comparison to the B73 Genome Sequence: A Characterization of Genotype-Specific Regions

Johann Joets, Clémentine Vitte and Alain Charcosset

Abstract

Maize is well known for its exceptional structural diversity, including copy number variants (CNVs) and presence/absence variants (PAVs), and there is growing evidence for the role of structural variation in maize adaptation. F2 is a European maize line resulting from a long-term independent evolution relative to the reference American line B73. It also presents strong heterosis when crossed to American lines related to B73 or PH207, which has been instrumental for the development of hybrid breeding in Northern Europe. De novo genome sequencing of the French F2 maize inbred line revealed 10,044 novel genomic regions larger than 1 kb, making up 88 MB of DNA, that are present in F2 but not in B73 (PAV). This set of maize PAV sequences allowed us to annotate PAV content and to identify 395 new genes. We showed that most of these genes display numerous features that suggest they are either rapidly evolving genes or lineage-specific genes. Using PAV genotyping on a collection of 25 temperate lines, we also analyzed and provided the first insights about PAV frequencies within maize genetic groups and linkage disequilibrium in PAVs and flanking regions. The pattern of linkage disequilibrium within PAVs strikingly differs from that of flanking regions and is in accordance with the intuition that PAVs may recombine less than other genomic regions. As it was shown by several other authors, most PAVs are ancient, while we show that some are found only in European Flint material, thus pinpointing structural features that may be at the origin of adaptive traits involved in the success of this material. We conclude by some words on future directions.

1.1 F2 Is Characteristic from a European Hybridization Event

The story of European maize traces back to its first introduction in 1493 by Columbus after his first trip to America. Being adapted to the tropical climate of the Caribbean, these varieties could be cultivated only in warm regions of the Mediterranean basin and would have been too late flowering to produce seeds in cooler environments. After this seminal trip, explorations lead to the rapid discovery of the northeast American coast, up to cool temperate climates of northern Canada. Most Native American people of the east coast and neighboring inland regions

J. Joets (✉) · C. Vitte · A. Charcosset
Genetique Quantitative et Evolution – Le Moulon, INRA, CNRS, AgroParisTech, Université Paris-Sud, Université Paris-Saclay, Gif-Sur-Yvette, France
e-mail: johann.joets@inra.fr

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_1

were relying to a large extent on the cultivation of specific maize varieties, referred to as Northern Flints because of their hard kernel texture. Their short planting to flowering interval was making them adapted to temperate environments. Genetic and historical investigations show that these temperate varieties were rapidly introduced into Europe and cultivated on a significant scale in Northern countries like Germany before 1539 (see Tenaillon and Charcosset 2011 for a review). Genetic analyses highlight that these two main introductions of maize into Europe at some step hybridized, possibly also with introductions of lesser importance, leading to varieties specific to mid-latitude European regions such as the Pyrenean valleys (Brandenbourg et al. 2017). Varieties from these introductions produced staple food in these regions until the late 1960s.

After WW2, traditional European varieties have been used to develop inbred lines, which were tested for their ability to produce hybrid varieties. Among these lines, F2 which stands for France n 2 was developed from the Lacaune population, cultivated on a cool South West France plateau at approximately 800–1000 meters elevation (Cauderon 2002). It proved outstanding in its ability to produce superior hybrids when crossed to inbred lines from North American origin, referred to as Dents because of their soft endosperm texture leading to a depression on the kernel crown. These first European Flints by American Dent hybrids were particularly successful for grain production in Northern Europe. This success can be interpreted as the combination of environmental adaptation features (adaptation to cool spring in particular) contributed by European Flints with yield potential contributed by American dents. Modern hybrid breeding for grain or silage production in North European regions is still based to a large extent on this pattern. As for F2 itself, it remained extremely successful and used in hybrids until the mid-1990s, especially when crossed to the American Dent lines PH207- or B73-related lines. Since that time, it has served as one of the three major progenitors of modern European Flint lines, along with lines F7 and Ep1.

Genotypic evaluations have confirmed a striking divergence between European Flint (i.e., F2) and American Dent lines (i.e., B73) (see Gouesnard et al. 2017). There are also striking phenotypic characteristics that differ between the two lines (Table 1.1). All elements therefore concur to expect large differences between the genomes of B73 and F2, possibly related to heterosis and adaptive traits.

1.2 B73- and F2-Specific Genome Region Discovery and Combination into a Draft B73–F2 Pan-Genome Sequence

Maize SV discovery at the whole-genome scale through comparative genomic hybridization arrays (aCGH)-based analysis of low copy regions led to detection of thousands of PAVs and CNVs between two American maize inbred lines (Springer et al. 2009; Beló et al. 2010). Probing of structural variation through a global analysis of read depth in over 100 maize lines showed that over 90% of the maize genome shows some degree of CNV between lines (Chia et al. 2012). While they allowed cost-effective and genome-wide discovery of PAVs/CNVs in multiple samples, these aCGH- and remapping-based studies did not allow discovering novel regions absent from B73. Discovery of over 2,000 new non-B73 genes was performed using massive mRNA sequencing on over 500 inbred lines, thus providing a cost-effective approach to solve this issue (Hirsch et al. 2014). Nevertheless, discovery of new genes with such mRNAseq-based strategy is dependent on sequencing depth and on the number of tissues and conditions analyzed. It is therefore likely to miss new genes with very low expression or expressed in very specific conditions. Moreover, this type of strategy does not provide sequence breakpoints, thus hampering exploration of underlying mechanisms, and is limited to analysis of the genic portion of the genome. Genome sequencing and de novo assembly can ultimately provide precise breakpoint positions, distinction between CNV and PAV, access to novel sequences, variant size information, and exploration of non-genic space. Targeted assembly of non-B73

Table 1.1 Summary of main phenotypic/adaptive differences between F2 and B73

Trait	F2	B73
Cold adaptation	Mid-tolerant	Sensitive
Leaf number[a]	14.1	20.6
Plant height (cm)[a]	142	210
Flowering time[a]	Early	Late (+20 days)
Kernel number/ear[a]	221	473
Endosperm	Hard (Flint)	Soft (Dent)

[a]Estimations from Bouchet et al. 2017

regions from elite Chinese and American lines led to the discovery of 5.4 MB of new sequence absent in the reference genome assembly (Lai et al. 2010). However, the low sequencing depth used (5X) limited the reconstruction of full-length PAV sequences. Because discovered PAVs were short and incomplete, complete annotation and anchoring to the reference genome were challenging, thus impeding functional prediction and breakpoint detection. Sequence assembly of the PH207 genome provided a matrix for reciprocal comparison of PH207 and B73 gene coverage using remapping of massive sequencing reads. It led to the discovery of over 2,500 genes, which were found specific to one genotype either partly or fully (Hirsch et al. 2016). However, analyses were focused on gene-annotated regions only, so this study did not identify the boundaries of the SVs containing these genes. In a complementary work, we produced a draft sequence assembly of the F2 genome and identified over 10,000 genomic regions present in F2 and absent from B73 (Darracq et al. 2018). New F2 regions make up 90 MB (4% of F2 genome size). Using RNAseq data from 12 tissues and conditions, we identified near 400 genes expressed in F2 PAVs. Expression breadth revealed that PAV genes are expressed in a limited set of conditions and at a lower rate than average B73 genes, consistent with previous results (Hirsch et al. 2016). Hence, while most F2-specific genes are likely present in our assembly (which covers 65% of the F2 genome), we likely did not explore enough conditions to have a RNAseq support for all new genes, and further transcriptome studies may help unravel more F2 specific genes.

Genome comparison studies provide a starting point to unravel the molecular origin and the function of maize structural variants. A consensus assembly that represents many individuals is likely to improve use of sequence-based chromatin and transcription data, as well as SNP detection. Decreasing the amount of spurious alignments would help to better estimating transcript abundance or heterozygosity prediction. How to best combine genomic sequences from several maize inbreds for aligning Illumina reads in a compute-efficient way remains a challenge (Consortium 2016; Hurgobin and Edwards 2017). While using each genome separately is an option, the rapid increase of whole-genome sequences will soon make it too computationally costly. Rather, we propose to build pan-genomic sequences by adding up the non-B73 genomic sequences to the B73 genome sequence. As a proof of concept, we built a first B73–F2 pan-genomic sequence, by adding up the 90 MB of F2-specific sequences to the 2.1 GB B73 genome sequence (Darracq et al. 2018). In the following sections, we will show how our approach can be used for studying (i) characteristics of PAVs and underlying genes, (ii) PAV LD properties, (iii) PAV history among maize inbreds, and (iv) perspectives for improved discovery and use in post-genomic studies.

1.3 F2 Non-B73 Genes Are Expressed in Other Maize Lines, but Are not Well Conserved Outside Maize

The 395 novel predicted genes present in F2 and absent of B73 are all supported by RNAseq experiments. In a comparison of RNAseq-based

abundance of F2 PAV genes versus B73–F2 shared genes, we showed that F2 PAV genes are expressed in less tissues than shared genes. This suggests that RNAseq-based identification of genes in F2 PAVs may have missed some genes due to lack of transcriptomic data in a large enough set of tissues/conditions. When comparing F2 PAV genes with transcriptome datasets from maize and related species, we showed that 90% have a blast best hit with a maize sequence, from another genotype other than B73, and only 8% have a best hit in closely related Poaceae species (20 in *Sorghum*, 3 in *Setaria*, 3 in *Saccharum*, 1 in *Miscanthus*, 1 in *Panicum*, 1 in *Tripsacum*, and 2 in *Oryza*). Interestingly, most orthologous sequences that were found derived from expressed sequence tags (ESTs) that were produced in the early 2000s for large numbers of maize genotypes, tissues, and conditions. By comparison, only 12 novel F2 sequences align to sequences from the pan-transcriptome assembled by Hirsch et al. (2014), which included more than 500 genotypes but from a single tissue. This suggests tissue/condition specificity of PAV gene expression and highlights the need for enlarging RNAseq datasets to improve discovery, annotation, and characterization of genotype-specific genes.

Functional annotation by search of sequence similarity with UniProtKB/Swiss-Prot proteins and InterPro protein domains allowed annotation of 91 F2 PAV genes. Among these, 17 (20%) are putatively involved in stress response and plant defense, 11 (12%) in biosynthetic processes, 10 (12%) in development, 5 (6%) in protein synthesis, and 5 (6%) in chromatin remodeling. For B73, PAV annotation was based on existing RefGen v2 5a annotation and provided a molecular functional prediction for 25 B73 PAV genes. Grouping of these molecular functions highlighted six sequences (25%) putatively involved in metabolism, four (16%) in stress response and plant defense, four (16%) in protein degradation, and two (8%) in cytoskeleton/microtubule. These results suggest that F2 PAV genes and B73 PAV genes are enriched in functions involved in stress response. Similarly, an enrichment of function related to stress response was observed in a set of maize PAV genes identified from a comparison between PH207 and B73 (Hirsch et al. 2016). Hence, transcriptome profiling in abiotic and biotic stress conditions is likely to greatly increase prediction and annotation of genotype-specific genes. Interestingly, in a recent study analyzing the diversity of 67 maize genomes from landrace representatives from the major maize genetic groups, including European lines, we uncovered that genes involved in abiotic stress tolerance have played a role in maize adaptation to European conditions (Brandenburg et al. 2017). This opens interesting perspectives in deciphering the role of PAVs in maize adaptation.

While this study allowed for prediction of PAV functions, protein prediction was successful for only 23% of the F2 novel genes sequences. This suggests that F2 PAV genes may be less conserved than other genes. To test this, we compared PAV and non-PAV genes in maize in terms of both number of genes with protein similarity, and levels of similarity to the protein sequence in an increasingly distant species set, from *Sorghum bicolor* to *Arabidopsis thaliana*. As predicted, the proportion of proteins with no significant similarity with other plant proteome is higher for F2 PAV genes (Fig. 1.1a), and when a protein is found, average identity is markedly (12 to 25%) lower for F2 PAV gene proteins than for B73 FGS proteins (Fig. 1.2b). This lower conservation suggests that PAV genes identified in F2 compared to B73 could have evolved more rapidly than non-PAV genes or emerged recently as novel genes.

With shorter size, shorter expression breadth, enrichment in stress-related functions, and lower conservation at the protein level than average genes, F2 PAV genes have many characteristics of orphan genes (Arendsee et al. 2014). Orphan genes either emerge de novo from non-genic sequence or derive from ancient gene duplications followed by divergent accumulation of mutations beyond recognition. Nevertheless, functional characterization of these genes is still challenging. Because discovery and annotation of PAV genes are a major goal in maize and plant biology, many laboratories are generating

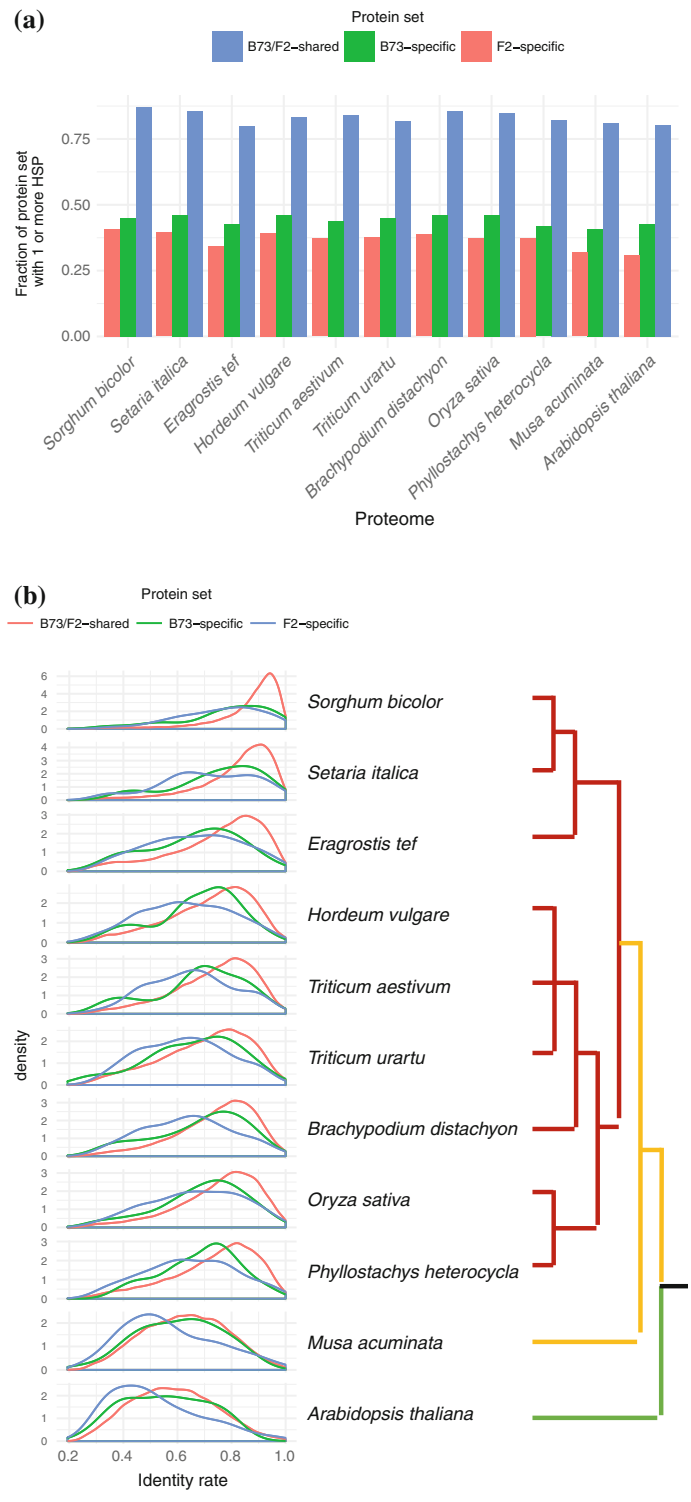

Fig. 1.1 Conservation of B73 and F2 presence/absence variation (PAV) proteins compared to B73–F2 shared proteins. **a** Fraction of protein sets (B73–F2 shared proteins, B73-present/F2-absent proteins, F2-present/B73-absent proteins) with at least one blastp hit (tilled HSP) (E value > = 10^{-3}) with several whole plant proteomes. **b** Distribution of identity rate of blastp best hit (tilled HSP) for the three protein sets against 11 whole plant proteomes. Plant proteomes are sorted according to the genetic distance with maize from sorghum to Arabidopsis, which is the most distant of maize. Length of branches of the phylogenic tree are arbitrary, red branches are for grasses, orange for monocot, and green for eudicot

RNAseq and proteome datasets to help in this task. We believe that this effort will provide important information for better understanding the origin and role of orphan genes.

On the other hand, it has been argued that most of the dispensable genes are members of duplicated gene or large gene family members (Swanson-Wagner et al. 2010). The absence of the gene could therefore be complemented by another member of the family. Of the 395 novels genes discovered in F2, only 116 exhibit greater than 50% identity over at least 80% of their length with a protein of B73, and therefore, 70% of these proteins have no or distant similarity with protein in B73. While this is certainly an underestimation of the number of unique PAV genes as the B73 and F2 genomes are not complete, it is possible that a significant fraction of PAV genes, and possibly biological functions, are absent in some genotypes.

1.4 The Dispensable Genome: A Genomic Faction that Recombines Less Than the Rest of the Genome

Linkage disequilibrium (LD) is the non-random association between alleles at different loci. LD contains information about recombination, demographic history, and gene conversion. LD between copy number variation and flanking SNPs has been found to be higher than between SNPs in genomic regions neighboring CNVs (Schrider and Hahn 2010). This was attributed to the fact that many CNVs have changed genomic location through recurrent duplications and deletions compared to other loci (Schrider and Hahn 2010). In the case of PAV, LD pattern between SNPs within the PAV or between a PAV and flanking marker should not follow these of CNVs. As presented above, PAV genes have a particular mutation pattern and this may impact local LD. But most PAVs do not harbor genes, and whether the whole PAV region evolves at a different rate than other loci remains to be elucidated. To get a first insight on LD pattern between PAVs and their flanking regions, we estimated LD extent for each PAV coded as 0 (absence)/1 (presence) or using the SNP located within the PAV and with shortest distance to the breakpoint. LD was then estimated between this reference polymorphism and SNPs of the flanking region, with increasing distance. While the first approach involves all individuals, for the second, LD can be estimated only when SNPs can be evidenced within the PAV, hence only in the subset of individuals that carry the present allele. For this, we developed a statistical approach to genotype PAV presence and absence alleles using low depth (3x–5x) resequencing data aligned on our B73–F2 pan-genome sequence and applied it on a dataset from a panel of 25 maize lines representing American and European maize genetic groups (Darracq et al. 2018). We compared these LD patterns with those estimated for reference genomic regions and their flanking regions. We showed that LD pattern between PAVs and their flanking regions resembles the same pattern observed between random genes and their flanking regions (Fig. 1.2a). While this might be due to our detection approach to discover PAVs, this first analysis shows that for these PAVs in our panel, LD decreases rapidly. This suggests that PAVs are likely not to be captured by genotyping SNPs, unless these are located within less than 1 kb of the PAV breakpoint.

To investigate whether PAVs recombine less than other genomic regions, we compared LD patterns within PAVs to LD patterns in their flanking regions. While LD depends on demographic history of the lines tested, this effect should be the same for two adjacent genomic regions such as a PAV and its flanking regions, thus giving a relative difference of local recombination rates. For this analysis, within-PAV LD was estimated by comparing pairwise SNPs located inside the variant sequence to pairwise SNPs in the PAV upstream or downstream flanking regions. On average, LD is stronger within a PAV as compared to the flanking regions (Fig. 1.2b). Hence, PAVs seem to recombine less than their flanking regions.

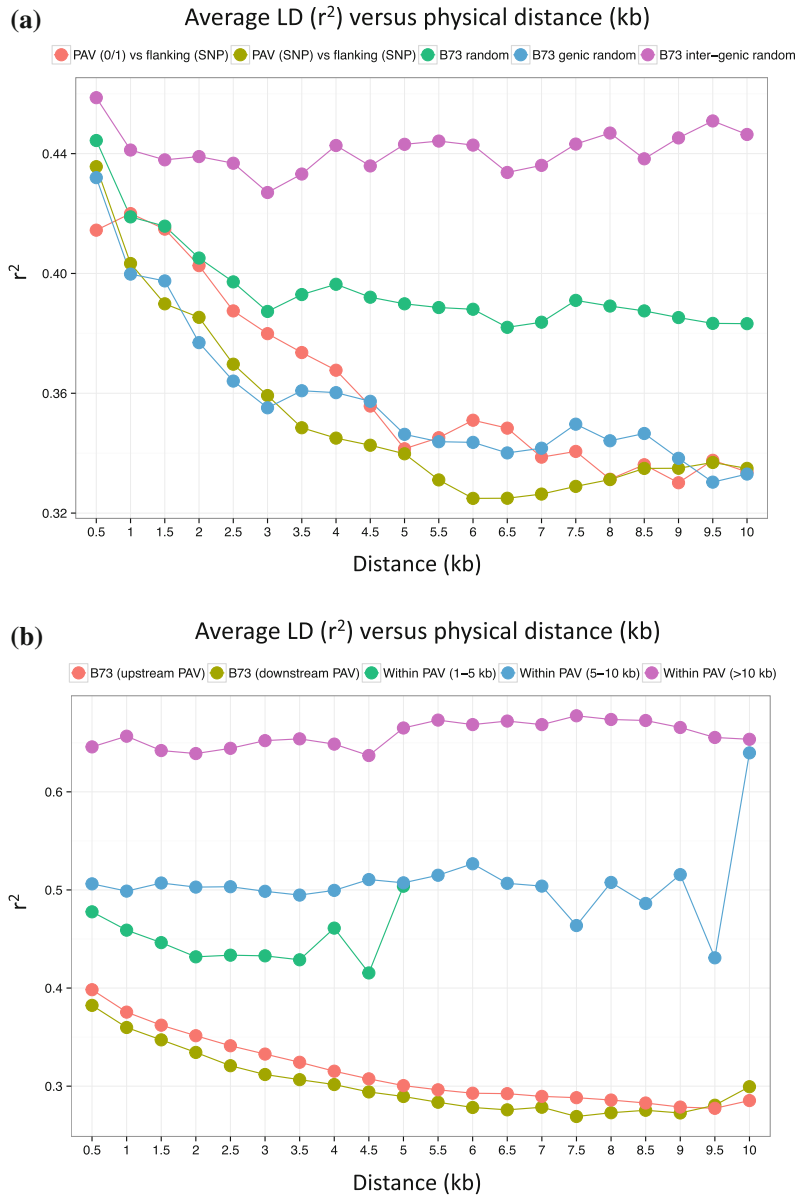

Fig. 1.2 Linkage disequilibrium (LD) decay pattern in presence/absence variation (PAV) regions. **a** LD decay between PAV and flanking region compared to LD decay between gene or TE and flanking regions, respectively. **b** Within-PAV LD compared to LD in flanking regions. PAVs were grouped into three classes according to their size; 1–5 kb PAVs green, 5–10 kb PAVs blue, >10 kb PAVs pink

This result may be due to the fact that PAV sequences can undergo recombination only when present in both gametes, a situation that is less frequent than for shared flanking regions. Of course, this situation depends on the PAV allele frequency, which also depends on the age of the PAVs, so we expect a large range of recombination rates among PAVs. Indeed, when considering PAVs individually, contrasting LD patterns can be observed. For instance, cases of very strong LD are found (Fig. 1.3 left), while in some cases LD patterns reveal subsets of recombining regions within the PAV sequence (Fig. 1.3 right). This difference is likely due to the differences in the date of appearance of the PAV in the population, its frequency in the population upon creation, as well as the temporal dynamics of this frequency.

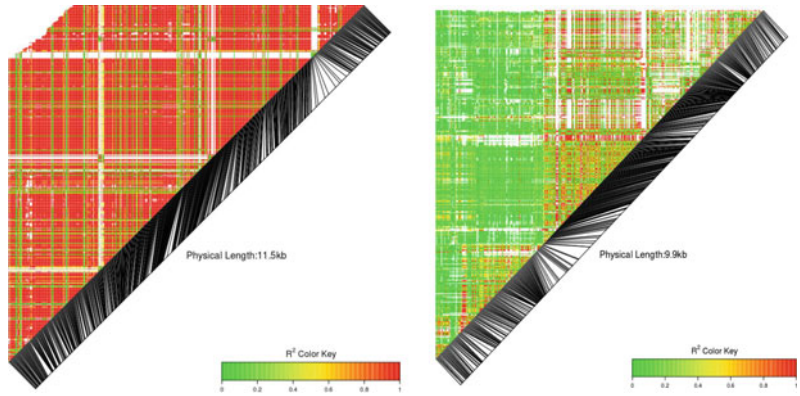

Fig. 1.3 Two examples of contrasted within presence/absence variation (PAV) linkage disequilibrium (LD) patterns. Left most of SNPs are in very high LD all along the PAV. Right two regions of high LD are separated by a breakpoint

1.5 Analysis of PAV Alleles at the Population Level

To investigate to what extent B73–F2 PAVs are conserved among maize genetic groups, we used the genotyping of PAV sequences in our temperate maize panel to estimate frequencies in the different genetic groups. As expected, F2 novel regions were more often present in other European Flints than in any other set of inbreds. Only a small number was detected in the Stiff Stalk group, to which B73 belongs and where the "absent" allele was found. Inbred lines from France or close proximity (Pyrenean) shared more variants with F2 than lines from any other origin, independent of their classification into European Flint and Northern Flint groups, thus reflecting the history of the European germplasm. PCA-based analyses from PAV or SNPs showed very similar classification, showing that SNPs and PAVs have segregated similarly.

A large proportion of PAVs are shared between F2 and at least one other European Flint or one Corn Belt Dent line, which were the most represented groups in our panel. PAVs that were found present in all the four genetic groups were also generally found at high frequency in all groups, suggesting an ancient and shared origin. Consistently, a comparative genomic hybridization experiments on 19 maize lines and 14 teosinte, the wild ancestor of maize, found that 86% of the SVs (CNV and PAV) that were identified were also present in teosinte (Swanson-Wagner et al. 2010). However, 347 PAVs were present only in maize but not in teosinte, and among them, 257 were present in only two to three maize lines suggesting these variants could be specific to maize. We also observed that when PAVs are present in only one genetic group their frequency is low in this group, suggesting the occurrence of recently emerged PAVs. Interestingly, among the 4,218 PAVs that we scored, 396 were found only in European Flints and 134 only in F2 (Darracq et al. 2018). Genotyping of these putative European-specific PAVs in larger maize panels will allow precise allele frequencies and group specificity to be determined.

1.6 Tomorrow's Challenges in Maize Structural Variation

Over the past decade, there has been a growing attention for structural variation in plant evolution. In maize, several genomic studies, including ours, have described some of the characteristics of CNVs and PAVs. But such studies are still in their infancy, and many questions remain to be solved. First, because the maize genome is highly repetitive, discovering structural variants in the repetitive fraction is still a challenge, and most structural variants that have been discovered are from low copy regions. Some studies have made

the choice to focus on genes, which is a cost-effective way of finding SVs with possible phenotypic impact (Hirsch et al. 2014). Using a non-targeted, without a priori approach, we could discover full-length PAVs containing both genic and non-genic regions and characterize their breakpoints. This gave us access to their full sequence content and made LD analyses possible. However, only a subset of our F2 PAVs could be anchored, either because their breakpoints could not be unambiguously anchored or because the assembly was not complete enough to extend PAVs to their biological breakpoints. In both, cases, these issues are linked to the highly repetitive nature of the maize genome, which impairs both unambiguous alignments of short reads in remapping experiments or in whole-genome assembly. This issue might soon be solved, as several maize whole-genome assemblies are under progress. High-quality metrics obtained from new assembly methodologies will open the way to whole-genome sequence comparison, thus eliminating the problem of aligning short reads. Such assemblies are now available for American lines (B73, PH207, W22, CML247) and European lines (EP1, F7). We will soon double this number by adding seven new genome sequences from lines of interest for the European community, and with contrasted genome sizes as well as the complete set of NAM founder parents.

A second challenge is to discover genes standing within these structural variants. As we presented, the particular features of these genes make them difficult to annotate, and the generation of large datasets of RNAseq and proteomic data in many tissues and conditions will be necessary to solve this problem. For this reason, for our seven genotypes and for B73, we are generating deep mRNAseq datasets from a set of tissues from standard- and abiotic-constrained conditions.

Once discovery and annotation of SV will be resolved, the next step will be to combine the information given by these new datasets to make the best use of it. Several laboratories are working on this question, and discussions are emerging. But this is only the beginning, and the maize community needs to organize.

Clearly, pan-genome sequence will be very useful for better analyzing phenotypic data at the molecular (methylome, transcriptome, proteome) or plant scale to find the underlying genetic components. Using the entire genomic information in GWAS will therefore be a major task in the coming years, and typing both SVs and SNPs will be necessary. We developed a pan-genome strategy that allows efficient alignment of resequencing data, as well as an efficient statistical methodology to classify PAVs as present or absent. This methodology can be used across a combination of a large number of maize lines. However, considering the history of maize, and the relatively limited bottleneck involved in its domestication, reconstructing haplotypes representing the entirety of maize genetic diversity will likely require retrieving information from hundreds of maize lines. This number is likely too high for producing public whole-genome sequence assembly resources for all of them, and defining a cost-effective strategy to do so will be an incoming task. We believe discussions at the community level will help build homogeneous datasets that can profit the whole community.

References

Arendsee ZW, Li L, Wurtele E (2014) Coming of age: orphan genes in plants. 19:698–708. https://doi.org/10.1016/j.tplants.2014.07.003

Beló A, Beatty MK, Hondred D et al (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120:355–367. https://doi.org/10.1007/s00122-009-1128-9

Bouchet S, Bertin P, Presterl T, Jamin P, Coubriche D, Gouesnard B, Laborde J, Charcosset A (2017) Association mapping for phenology and plant architecture in maize shows higher power for developmental traits compared with growth influenced traits. Heredity 118:249–259

Brandenburg J-TT, Mary-Huard T, Rigaill G et al (2017) Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLoS Genet 13:e1006666. https://doi.org/10.1371/journal.pgen.1006666

Chia J-MM, Song C, Bradbury PJ et al (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807. https://doi.org/10.1038/ng.2313

Cauderon A (INRA, ed) (2002) L'INRA dans l'amélioration des plantes des «Trente Glorieuses» à la lumière des préoccupations actuelles, in: L'Amélioration des Plantes, continuités et ruptures (http://www.inra.fr/gap/viescientifique/animation/colloqueAP2002/Cauderon.pdf)

Consortium CP-G (2016) Computational pan-genomics: status, promises and challenges. Brief Bioinform. https://doi.org/10.1093/bib/bbw089

Darracq A, Vitte C, Nicolas S, Duarte J, Pichon JP, Mary-Huard T, Chevalier C, Bérard A, Le Paslier MC, Rogowsky P, Charcosset A, Joets J (2018) Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. BMC Genom 19:119

Gouesnard B, Negro S, Laffray A, Glaubitz J, Melchinger A, Revilla P, Moreno-Gonzalez J, Madur D, Combes V, Tollon-Cordet C, Laborde J, Kermarrec D, Bauland C, Moreau L, Charcosset A, Nicolas S (2017) Genotyping-by-sequencing highlights original diversity patterns within a European collection of 1191 maize flint lines, as compared to the maize USDA genebank. Theor Appl Genet 130: 2165–2189

Hirsch C, Hirsch CD, Brohammer AB, et al (2016) Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. Plant Cell 28:tpc.00353.2016. https://doi.org/10.1105/tpc.16.00353

Hirsch CN, Foerster JM, Johnson JM, et al (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell Online tpc.113.119982. https://doi.org/10.1105/tpc.113.119982

Hurgobin B, Edwards D (2017) SNP discovery using a pangenome: has the single reference approach become obsolete? Biology 6:21. https://doi.org/10.3390/biology6010021

Lai J, Li R, Xu X et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:1027–1030. https://doi.org/10.1038/ng.684

Schrider DR, Hahn MW (2010) Lower linkage disequilibrium at CNVs is due to both recurrent mutation and transposing duplications. Mol Biol Evol 27:103–111. https://doi.org/10.1093/molbev/msp210

Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez LA, Barbazuk A, Jeddeloh JA, Nettleton D, Schnable PS, Ecker JR (2009) Maize inbreds exhibit high levels of copy number variation (cnv) and presence/absence variation (pav) in genome content. PLoS Gen 5(11): e1000734

Swanson-Wagner RA, Eichten SR, Kumari S et al (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20:1689–1699. https://doi.org/10.1101/gr.109165.110

Tenaillon MI, Charcosset A (2011) A European perspective on maize history. Comp Rendus Biol 334(3): 221–228

The Maize Pan-Genome

2

Alex B. Brohammer, Thomas J. Y. Kono and Candice N. Hirsch

Abstract

The pan-genome of a species is comprised of genes/sequences that are present in all individuals in the species (core genome) and genes/sequences that are present in only a subset of individuals within the species (dispensable genome). In maize, the study of the pan-genome began in the 1940s through cytogenetic experiments and has seen an increased focus in research over the last decade largely driven by advances in genome sequencing technologies. It is estimated there are at least 1.5x as many genes in the pan-genome (greater than 60,000 genes) as there are in any individual's genome (~40,000 genes), with even more variation outside the gene space being observed. This variation has been associated with phenotypic variation and is hypothesized to be an important contributor to the high levels of heterosis often observed in maize hybrids. Due to the high level of variation and the existing genetic and genomic resources, maize has become a model species for plant pan-genomics studies. This chapter will review the mechanisms that can create genome content variation, tools that are available to study the pan-genome, the history of maize pan-genome research ranging from the early cytogenetic studies to today's genomics-based approaches, and the functional consequences of this variation.

2.1 Introduction

By definition, the pan-genome refers to the non-redundant set of sequences distributed throughout the population of a particular species. A pan-genome consists of two sets of sequences: those present in every individual in the population, the core genome, and those present in only a subset of individuals, the dispensable genome. The dispensable genome can be further partitioned based on a frequency spectrum. Genes present in low frequencies are part of the "cloud" set, while those in intermediate and high frequencies are part of the "shell" and "soft core" sets, respectively (Koonin and Wolf 2008).

The concept of a pan-genome was introduced by the bacterial community to describe the extensive variation in genome content between species (Tettelin et al. 2005; Medini et al. 2005; Hogg et al. 2007; Tettelin et al. 2008). Technological advances and reduced sequencing technology costs have permitted the pan-genome concept to be extended beyond bacterial species to the plant and animal kingdoms (Li et al. 2010; Computational Pan-Genomics Consortium

A. B. Brohammer · T. J. Y. Kono · C. N. Hirsch (✉)
Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul 55108, MN, USA
e-mail: cnhirsch@umn.edu

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_2

2016). Within the plant kingdom, pan-genome analyses have been applied to a number of model and crop species such as *Arabidopsis thaliana* (Cao et al. 2011; 1001 Genomes Consortium 2016), *Brachypodium distachyon* (Gordon et al. 2017), *Brassica oleracea* (Golicz et al. 2016), *Glycine soja* (Li et al. 2014), maize (*Zea mays*; Hirsch et al. 2014), *Medicago truncatula* (Zhou et al. 2017), *Oryza sativa* (Yao et al. 2015), soybean (*Glycine max*; Anderson et al. 2014), and wheat (*Triticum aestivum*; Montenegro et al. 2017).

Depending on the number of genomes that need to be surveyed to capture the full suite of dispensable genes in a species, a pan-genome can be considered open or restricted. The former is common of bacterial species, where with each additional genome that is sequenced new genes are added to the species pan-genome (Tettelin et al. 2008). In contrast, restricted genomes like maize are typical of plant and animal species, where the majority of the pan-genome is captured in a relatively limited set of individuals. In maize, through a transcriptome-based analysis it was estimated that approximately 350 lines were needed to capture the suite of dispensable genes transcribed in the seedling (Hirsch et al. 2014).

Genome content variation in pan-genomes is often described in the context of gene copy number variation (CNV) and gene presence/absence variation (PAV). Copy number variation describes the situation in which additional copies of a particular gene exist in one individual compared to another, and PAV is simply the extreme form of CNV, where one individual possesses one or more copies and another has zero copies of the gene. Genome content variants can result from recombination-based mechanisms, replication-based mechanisms, or other molecular mechanisms and can be divided into two broad categories based on whether they lead to a balanced or unbalanced outcome. This chapter will expand on these mechanisms that generate genome content variation in plant pan-genomes, tools to measure genome content variation, historical and contemporary knowledge on the maize pan-genome, and the functional importance of this variation in driving phenotypic variation within the species.

2.2 Mechanisms that Generate Genome Content Variation

2.2.1 Transposable Elements

Transposable elements (TEs) are genomic elements that have the ability to move in the genome either through a copy-and-paste or cut-and-paste mechanism. Transposable elements were first identified by Barbara McClintock through studying disruption of pigments in maize kernels (McClintock 1950) and comprise approximately 85% of the maize genome (Schnable et al. 2009). In addition to having direct effects on protein-coding sequence and transcript regulation (Tenaillon et al. 2010), TEs also provide multiple avenues for generation of genome content variation. Some classes of TEs "capture" and shuffle gene fragments or entire genes during transposition such as Pack-MULEs and *Helitrons*. Additionally, TEs are a form of dispersed homologous sequence throughout the genome, which can lead to ectopic recombination and the generation of novel gene sequences (Bennetzen and Wang 2014). Finally, the presence of TEs can stimulate meiotic recombination, presumably through the generation of transposase-induced double-strand breaks (Yandeau-Nelson et al. 2005). Subsequent error-prone repair of these breaks then provides further opportunity for genome content variation.

2.2.2 Unequal Recombination

Unequal recombination occurs when homologous chromosomes do not pair exactly during meiosis, and recombination results in gametes with differing DNA content. This is particularly prone to occur in regions of the genome that are already duplicated, because paired sequences may be locally homologous, but may not be globally homologous. Recombination between these improperly paired chromosomes then

generates some gametes with more DNA than the progenitor cell, and some gametes with less DNA. Genes arranged in tandem duplicate arrays are common in maize (Messing et al. 2004; Schnable et al. 2009) and provide opportunities for genome content variation via unequal pairing and recombination of duplicated sequences. For example, the *A1-b* locus in maize is a naturally occurring tandem duplication of the anthocyaninless1 (*a1*) gene that has been well characterized for unequal recombination (Yandeau-Nelson et al. 2006). In this case, unequal pairing of the duplicated genes occurred preferentially between homologous chromosomes, but could also occur between sister chromatids. Unequal recombination rates at the duplicated locus were similar to equal recombination rates at non-duplicated *a1* loci, suggesting that unequal recombination is a common phenomenon at this locus.

2.2.3 Non-allelic Homologues

Similarly to unequal recombination, segregation of single-copy homologues in non-allelic positions can also lead to changes in gene copy number in the genome (Emrich et al. 2007). Mating between two individuals carrying single-copy homologues in non-allelic positions will result in progeny that are hemizygous for each of the homologues. Independent assortment, or meiotic recombination if the homologues are physically linked, generates gametes that have variable copy number for the homologues. Inbred progeny produced from these gametes then have zero, one, or two copies of the non-allelic homologues, resulting in apparent *de novo* copy number variation. An example of this phenomenon in maize is two loci involved in elongation of fatty acid precursors for surface lipids, gl8a and gl8b. These two loci are unlinked paralogs with 96% nucleotide sequence identity in B73 that can form *de novo* copy number variation (Dietrich et al. 2005). On a genome-wide scale, several dozen genes were documented to be non-allelic homologues in a single recombinant inbred line population that showed apparent *de novo* copy number variation through segregation of the non-allelic homologues (Liu et al. 2012). This *de novo* copy number variation was hypothesized to contribute to the phenotypic transgressive segregation observed in the population across a number of phenotypic traits.

2.2.4 Horizontal Gene Transfer

Horizontal gene transfer (HGT) refers to the asexual transfer of genes between organisms of divergent evolutionary lineages. Maintenance of a newly transferred gene as a segregating genome content variant depends on several events. First, the horizontally transferred gene must integrate into a cell that gives rise to gametes in order for it to be transmitted into subsequent generations. It must then not be lost due to genetic drift and provide strong enough selective advantage to be maintained in a population. As such, it is hypothesized that horizontally transferred genes that persist as segregating variation within a population have a particularly high likelihood of contributing to phenotyping variation.

Horizontal gene transfer was first observed in bacteria (Freeman 1951) and is now known to be highly prevalent among bacterial species. In bacteria, HGT occurs through random uptake of extracellular DNA, incorporation of viral DNA into the host genome, or direct transfer of plasmids among individuals (Syvanen 2012). While rare in plants, HGT has been observed via viral DNA repeats in *Nicotiana tabacum* (Bejarano et al. 1996). Expressed transfer DNAs from *Agrobacterium rhizogenes* have also been observed in cultivated sweet potato (Kyndt et al. 2015). Plant-to-plant HGT has also been documented in parasitic species. For example, a nuclear gene in *Striga hermonthica*, a hemiparasitic plant that can cause devastating crop loss in species such as *Sorghum bicolor*, has been found to have high similarity to genes from *S. bicolor*, suggesting HGT as an origin for this gene in *S. hermonthica* (Yoshida et al. 2010).

2.2.5 Genome Duplication and Fractionation

When a genome undergoes a whole genome duplication event, it generates four copies of each nuclear gene where there were previously just two. New mutations can then begin to cause the function of the duplicates to diverge. Under classical models, the net direction of molecular evolution will be toward the ancestral state of two functional copies of each gene. Three major paths to this outcome are that one duplicate evolves a new function (Ohno 1970), the copies are retained and each partially loses function (Force et al. 1999), or one of the copies completely loses function (Jacq et al. 1977). Following a whole genome duplication, the most common mechanism to restore the ancestral diploid function is through fractionation (Langham et al. 2004; Tang et al. 2008).

An ancient genome duplication event in the ancestor of maize resulted in two subgenomes in present-day maize. Analysis of the B73 reference genome assembly showed that one subgenome has greater gene retention than the other, and these subgenomes were named "Maize1" and "Maize2," respectively (Schnable et al. 2011). The paralogs lost during fractionation are not completely consistent between individuals within the species and this variation in gene loss during fractionation generates genome content variation within the species (Brohammer et al. 2018). Many genes that show presence/absence variation within maize also show sequence similarity to genes in closely related grass species (Hansey et al. 2012; Hirsch et al. 2014). This suggests that these genes were present before the divergence of the maize lineage from other grass species and were differentially lost among maize individuals.

2.3 Contemporary Tools to Measure Genome Content Variation

2.3.1 Reference-Based Methods

Reference-based methods used to measure genome content variation within species include oligonucleotide arrays and next-generation sequencing (NGS) read mapping. Oligonucleotide arrays were the first reference-based method used for conducting genome-wide surveys of genome content variation within maize (Springer et al. 2009; Beló et al. 2010). A specific technique called array-based comparative genomic hybridization (aCGH) was particularly important to advancing our knowledge of PAV and CNV in maize. In this method, two labeled DNA samples are hybridized to probe sequences designed to target regions throughout the genome, and signal intensity from each labeled sample indicates its relative copy number. A major limitation to aCGH, and arrays in general, is the inability to detect sequences absent from the reference genome since probes are often designed from a single reference individual. Related issues brought about by limitations of probe design from a single reference individual include biased CNV detection toward deletion discovery and a reduced ability to evaluate regions of high sequence diversity.

Unlike aCGH, NGS methods allow for the discovery of the full suite of structural variants within the species including sequences outside the reference genome (Young et al. 2016). There are three common NGS structural variant detection methods: read depth, split read, and read pair. The read-depth method relies on sequence read depth from mapping reads to a reference genome assembly as a proxy for copy number. Both the split-read and read-pair methods take advantage of imperfect mapping to identify genomic rearrangements and allow for the detection of all structural variant classes, including inversions and translocations. Paired-end and mate-pair sequence reads have an expected insert size between the two sets of reads. Deviation from these expected distances between the two reads can be used to identify structural variations. The read-pair method uses reads whose distance or orientation between mapped reads from the same fragment is discordant with the reference genome to detect structural variation. The split-read approach to structural variation detection uses information from paired-end sequence reads where one of the

pairs maps accurately while the other pair maps only partially or fails to map entirely. The split-read approach can also be expanded to splitting an individual read and identifying reads in which only a portion of the read can accurately map to the reference genome as another method to identify structural variation.

Each method of NGS structural variation detection has its own set of biases (Alkan et al. 2011), and each has variable sensitivities. Many of the available structural variation callers were originally developed to work with human cancer data or model mammalian species and may provide unreliable results or require extensive knowledge and tuning of parameters to be properly used with plant genomes. Combining at least two of these structural variation detection methods into a hybrid structural variation caller (i.e., SURVIVOR; Jeffares et al. 2017) that reports consensus structural variations can overcome some of these issues. Additionally, some of these methods rely on imperfect read mapping, which can be prevalent when mapping short NGS reads to highly repetitive plant genomes even in the case of reference genome reads mapping to the reference genome assembly. Increased read coverage and optimization of mate-pair library sizes can mitigate this challenge; however, long-read sequencing technologies offer the most promise for avoiding inconsistent structural variation detection in repetitive regions and for the detection of large structural variants.

2.3.2 Non-Reference-Based Methods

With reference-based variant detection, there is an ascertainment bias that is caused by the reliance on a single reference genome assembly. One method for characterizing gene content variation beyond a single reference genome assembly is through direct comparison of multiple *de novo* genome assemblies. Schatz et al. demonstrated the power of this approach by generating *de novo* genome assemblies of *indica*, *aus*, and temperate *japonica* rice strains, where they identified several megabases of variable sequence between the three strains (Schatz et al. 2014). This approach has also been used in maize where approximately thousands of novel genes were identified in a comparison of *de novo* genome assemblies of elite inbred lines from opposite heterotic groups (Hirsch et al. 2016; Darracq et al. 2018).

Direct comparison of whole genome *de novo* assemblies allows for detailed analysis of variation outside of a single reference genome; however, a major disadvantage is the cost and computational effort required to bring these studies to fruition. This disadvantage is important for pan-genome studies because it often leads to a small number of genotypes being assayed and an underestimate of dispensable genome content within species. An alternative approach is to use the transcriptome as a proxy to evaluate the gene space within a species pan-genome. This approach has the advantage of reducing both the amount of sequencing and computation required in pan-genome studies. In maize, the gene space is only ~97 MB of the genome, and as such, this approach was able to be used to study the maize pan-genome using over 500 accessions (Hirsch et al. 2014).

Recent improvements in assembly algorithms and the continued decline in sequencing costs are making multiple *de novo* genome assemblies within a species more practical (Schatz et al. 2014; Wetterstrand 2018). An example of this shift toward the generation of *de novo* genome assemblies for pan-genome analysis is the assembly and annotation of a panel of 54 *Brachypodium distachyon* accessions by Gordon and colleagues (Gordon et al. 2017). For seven years, only two reference genome assemblies for maize were available: the B73 reference genome, and Palomero Toluqueño, a popcorn landrace (Vielle-Calzada et al. 2009). In the span of just three years, nine additional genome assemblies were made publicly available (W22—GenBank assembly accession GCA_001644905.2; F7 and Ep1—(Unterseer et al. 2017); PH207—(Hirsch et al. 2016); B73—(Jiao et al. 2017); F2—(Darracq et al. 2018); Mo17, B104, and CML247 (Maize Genetics and Genomics Database 2017)).

New and emerging technologies that provide long-range information will help to further

improve genome assembly and facilitate structural variant discovery. This information can come from special library preparation protocols for short-read sequencing, long-read sequencing, or large-scale optical maps. For example, 10x Genomics linked-reads are synthetic long reads that preserve single-molecule information through microfluidic encapsulation technologies. This technology is similar to Illumina TruSeq Synthetic Long-reads (formerly Moleculo), but does not attempt to reconstruct each fragment. The Dovetail Chicago library preparation protocol relies on the Hi-C method of cross-linking DNA to capture long-range information, and like the 10x Genomics method, the processed reads can be read-out by a short-read sequencer such as an Illumina HiSeq. Third-generation single-molecule sequencing, which includes the technologies of Pacific Biosciences Inc. and Oxford Nanopore Technologies, sequence long DNA fragments to provide long-range linkage information. Finally, a separate method of preserving long-range information is through the construction of optical maps (i.e., OpGen and BioNano Genomics), which use restriction sites as "fingerprints" to resolve chimeric assemblies and identify large structural variations.

2.3.3 Iterative Mapping and Assembly

A common approach to querying population-scale variation in plant pan-genomes is iterative mapping and assembly. An example of this approach was recently published by Yao et al. who analyzed 1,483 cultivated rice accessions to identify non-reference genome assembly sequences (Yao et al. 2015). In this strategy, all of the individuals were sequenced at low coverage and then aligned to the reference genome. After filtering to remove contaminants and low-quality reads, the unmapped reads represent dispensable genome sequence. Yao et al. assembled the unmapped reads from *indica* and *japonica* separately so that the dispensable genome of each subspecies could be studied. After annotating protein-coding genes and transposable elements in each dispensable genome, they determined the genomic positions of ~80% of these features relative to the Nipponbare reference genome using linkage disequilibrium mapping. The iterative map and assemble approach allows for a larger portion of the natural variation to be sampled at a relatively low cost compared to *de novo* assemblies. A limitation of the method is that the specific breakpoints of the PAV are often not clear.

2.4 History of Maize Genome Content Variation Studies

Over the 9 years that have passed since the original publication of the B73 reference assembly (Schnable et al. 2009), the maize community has developed a nuanced understanding of genomic variation, in particular structural variation within the species. Maize genome content studies can be reviewed as a progression through four relatively distinct epochs: molecular and cytogenetic studies of large-scale chromosomal aberrations, Sanger sequencing applied to bacterial artificial chromosomes (BACs), whole genome-scale studies using array technologies, and application of next-generation sequencing to study genome content variation across numerous genotypes. These eras represent a timeline that spans nearly 70 years, with a number of seminal discoveries made during each era (Fig. 2.1).

2.4.1 Molecular and Cytogenetic Era

The study of structural variation in maize can be traced back to early observations of genome-size variation among maize and its wild relatives. Extraordinary levels of variation for nuclear DNA content were observed between different maize inbreds and landraces ranging from 9.4 to 25.2 pg 4C content values (Laurie and Bennett 1985). Much of this variation in genome size was attributed to the presence of supernumerary B chromosomes (Ayonoadu and Rees 1971; Poggio et al. 1998), and variation in heterochromatic

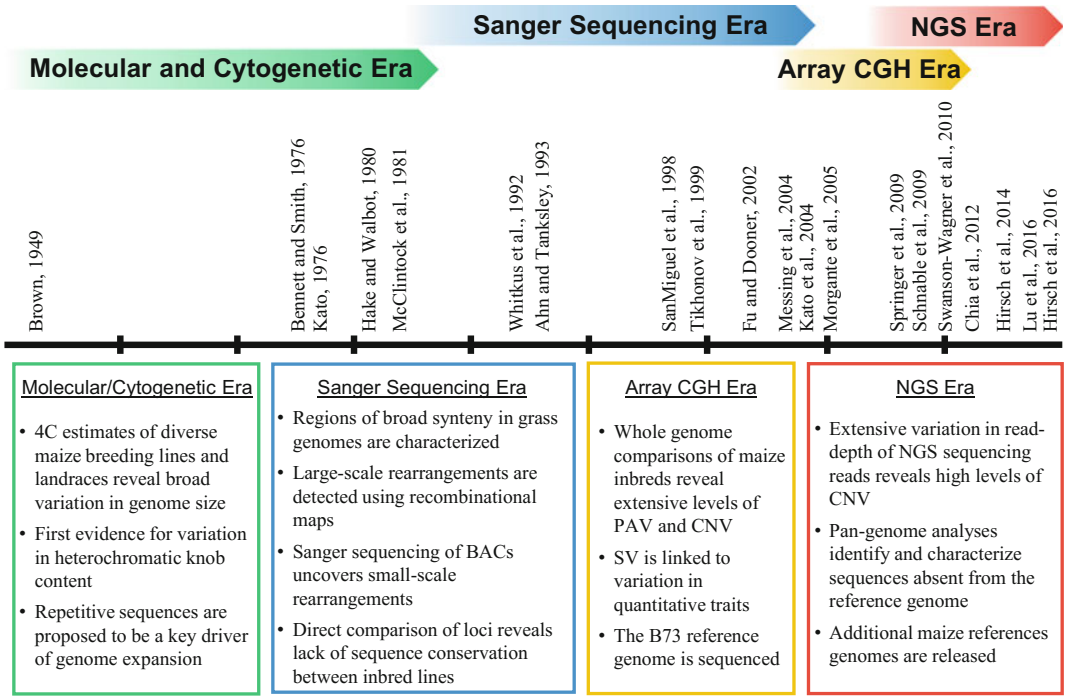

Fig. 2.1 Timeline of seminal studies leading to our current understanding of the maize pan-genome and functional consequences of genome content variation within maize. BAC—bacterial artificial chromosome; PAV—presence/absence variation; CNV—copy number variation; SV—structural variation; NGS—next-generation sequencing

knob content that makes up over 8% of the genome on average (Brown 1949; Kato 1976; McClintock et al. 1981; Peacock and Dennis 1981; Rayburn et al. 1985). Wide variation in the copy number of repeat sequences has also been widely observed in maize using molecular and cytogenetic approaches. These high-repeat sequences included ribosomal DNA (rDNA) repeats (Phillips et al. 1974; Buescher et al. 1984), centromere satellite DNA repeats (CentC) (Albert et al. 2010), telomere repeats (Burr et al. 1992), and dispersed repetitive sequences (Hake and Walbot 1980; Flavell 1986; Rivin et al. 1986; SanMiguel and Bennetzen 1998; Meyers et al. 2001). More recent surveys of the maize genome using modern cytogenetic and genomics techniques have confirmed these findings regarding variation in repetitive DNA content between maize lines (Kato et al. 2004; Liu et al. 2017).

2.4.2 Sanger Sequencing Era

The standardization of shotgun sequencing improved protocols for BAC library construction, and development of bioinformatic algorithms gave rise to the next era in the study of maize genome content variation in the late 1990s and early 2000s. Comparisons of orthologous regions between related grasses using recombination maps generally revealed broad synteny (Whitkus et al. 1992; Ahn and Tanksley 1993); however, in some cases large-scale rearrangements were observed (reviewed in Gale and Devos 1998). Subsequently, sequencing-based analyses of classical loci showed that smaller-scale rearrangements of orthologous sequence were much more common (Tikhonov et al. 1999). Soon thereafter, a landmark study discovered that the variation seen between orthologous regions could also be found between

maize inbred lines. Using the inbred lines McC and B73 to examine sequence variation at the *bz* locus, it was shown that four of the predicted genes in the McC haplotype were absent from B73 and many of the retroelements present were derived from independent insertion events (Fu and Dooner 2002). To determine if this result was due to a peculiarity between McC and B73, the region was evaluated across 10 separate inbred lines and four distinct structural variation haplotypes were found. In an accompanying commentary, it was hypothesized that the PAV between haplotypes was the result of differential fractionation between McC and B73 (Bennetzen and Ramakrishna 2002). The *z1C-1* locus was also evaluated using Sanger sequencing and significant variation in gene collinearity between the B73 and BSSS53 haplotypes was observed (Song and Messing 2003). A larger-scale comparison of 2.8 Mb of sequence between B73 and Mo17 revealed extensive stretches of nonhomology, in which more than one-third of the genes in the regions examined were variable in their presence (Brunner et al. 2005).

These studies raised numerous questions. *What is the genetic mechanism that gives rise to these presence/absence variants? What proportion of the gene complement is dispensable? Do presence/absence variants encode functional proteins?* The first of these questions was explored in a follow-up study by Dooner and colleagues who found that the variability in genic content at the *bz* locus could be attributed to *Helitron* elements (Lai et al. 2005). This was further supported via a genome-wide comparison of the inbred lines, B73 and Mo17, in which it was estimated that only ~80% of genomic segments were shared between these two lines based on hybridization to probes designed from genic sequences (Morgante et al. 2005). In-depth characterization of nine of the non-shared sequences showed that all but one displayed the hallmarks of *Helitron* capture (Morgante et al. 2005). At this time, prior to the completion of the B73 reference genome, it was hypothesized that any one line would contain only ~85% of functional maize genes (Buckler et al. 2006).

2.4.3 Array-Based Comparative Genomic Hybridization Era

The question of how many maize genes are affected by structural variation genome-wide was not addressed until the publication of the B73 reference genome (Schnable et al. 2009) and the subsequent development of an aCGH platform (Springer et al. 2009). A seminal paper from this era by Springer et al. showed that 180 high-confidence genes were present in B73 and absent in Mo17 (Springer et al. 2009). In addition to over 400 CNVs, a 2.6 MB stretch of sequence harboring 31 genes was identified that was completely missing from 17 of 24 inbred lines that were subsequently evaluated. This pattern of CNVs being common in maize populations has been recapitulated in other studies. A comparison of 14 inbred lines showed approximately half of over 2,100 identified CNVs were at high allele frequency (Beló et al. 2010). In a further comparison of 19 diverse maize inbred lines and 14 teosinte accessions, 3,410 CNVs were detected, ~86% of which were shared between maize and teosinte (Swanson-Wagner et al. 2010). These studies marked an important advance in knowledge not only due to the genome-wide scale of the studies, but also because they showed that low-copy expressed genes can be PAVs and CNVs, not just repetitive elements and pseudogenes.

2.4.4 Second- and Third-Generation Sequencing Era

The growth of next-generation sequencing technologies is closely tied to the next era of maize genome content variation studies. The initial maize HapMap study utilized sequencing-by-synthesis technology to inventory variation in the low-copy portion of the genome across 27 diverse inbred lines and estimated that B73 contained only 70% of the low-copy maize sequence (Gore et al. 2009). The second-generation HapMap study also

inventoried standing variation, but in an expanded collection of 103 inbred lines that included landraces and wild relatives (Chia et al. 2012). This study described the maize genome as being in "flux" with high levels of read-depth variants (RDVs). This description was based on scanning the genome in 10-kb bins and finding that more than 90% of the tested bins displayed greater than twofold variation in read depth across the individuals. Further, these RDVs were enriched for GWAS hits indicating their importance to phenotypic variation.

A number of subsequent studies have expanded beyond the reference genome assembly using iterative mapping and assembly approaches. In the first of this type, a set of six elite Chinese inbred lines were resequenced, and 570 novel gene sequences absent from the B73 assembly with an average coding sequence length of 527 bp were discovered (Lai et al. 2010). Of these 570 novel genes, 413 had high coverage from B73 resequencing reads while the remaining 157 did not, suggesting that the latter were true PAVs. Further analysis of the subset of the PAVs that did not have high resequencing coverage showed that many segregated in accordance with the heterotic group and did not have paralogs elsewhere in the genome. A similar approach was taken using RNA-seq of 21 diverse inbred lines across heterotic groups that identified 1,321 novel transcripts outside of the reference genome assembly, in which ~11% were heterotic group specific (Hansey et al. 2012). Finally, in a study of 503 diverse inbred lines that again used an RNA-seq mapping and assembly approach, over 20,000 transcribed sequences were identified that were not present in the B73 reference genome assembly, and it was determined that in this set of lines the closed maize pan-genome could be represented by ~350 lines (Hirsch et al. 2014). Subsequently, a novel method to convert GBS tags to pan-genome anchors across more than 14,000 genotypes found that B73 represents ~74% of the low-copy sequence present in maize (Lu et al. 2015). In this study, PAV SNPs were enriched for significant GWAS hits, but they were also negatively correlated with gene density and recombination frequency.

A new era in the study of maize genome content variation is emerging with the publication of multiple *de novo* genome assemblies and the availability of a new B73 reference genome assembly. The new B73 reference genome is a substantial improvement over the previous Sanger-based assembly with a 52-fold increase in contig length. Comparisons of this B73 genome assembly with the optical maps of two other inbreds, Ki11 and W22, showed that only 32% and 39% of the optical maps could be mapped to B73, respectively. Moreover, a large proportion of the aligned region showed evidence for structural variation including 257 PAVs missing in Ki11 and W22 (Jiao et al. 2017). *De novo* assembly of Iodent founder line PH207 allowed for a direct genome to genome comparison of gene content to B73 and reported 1,169 B73- and 1,545 PH207-specific genes in addition to extensive variation in gene family size (Hirsch et al. 2016). F2, an important inbred line in France, was assembled and 88 Mb of sequence was reported as unique to F2 in a comparison to B73 (Darracq et al. 2018).

2.5 Functional Importance of Genome Content Variation

2.5.1 Gene and Genome Evolution

Genome content variation represents an important class of potentially functional genetic variation. Duplication or deletion of genomic regions may have strong impacts on phenotypic variation, presumably because they disrupt the stoichiometry of gene products in physiological contexts (Torres et al. 2008). This disruption, however, is not necessarily detrimental. In the short term, changes in genome content may confer resilience to sudden stresses (Yona et al. 2012). In longer terms, changes in genome content may provide the starting point for evolutionary novelty and species diversification (reviewed in van de Peer et al. 2017).

Considering a single genetic locus, individuals that contain a gene (or multiple copies of a gene) that is not in the genome of others in a population may be able to perform unique biochemical functions, which may then increase variation for fitness. This is a major mechanism underlying the rise and spread of resistance to certain biotic (Cook et al. 2012) or abiotic (Maron et al. 2013) stresses. Duplicated genes may also provide a starting point for the evolution of novel gene function, because one copy of the gene is potentially released from purifying selection, allowing it to diverge in function (Ohno 1970; Näsvall et al. 2012). Genome content variants outside of protein-coding sequences may also have phenotypic effects and, thus, contribute to fitness variation. For example, maize transposable elements have been shown to influence neighboring gene expression, resulting in alteration of plant morphology (Studer et al. 2011), and abiotic stress response (Makarevitch et al. 2015). However, maintenance of increased copy number or unique biochemical pathways come at a cost, and gene duplicates are often purged in the absence of selective pressure to maintain them (Berglund et al. 2012).

2.5.2 Phenotypic Association and Cloned Genes

The phenotypic importance of genome content variation (CNVs and PAVs) has been shown through a number of genome-wide studies. The second-generation maize HapMap study (Chia et al. 2012) was particularly noteworthy as one of the first genome-wide studies to relate genome content variation to phenotypic variation in traits of agronomic importance. A subsequent association mapping experiment incorporated data from the HapMap studies to perform association mapping across 41 diverse phenotypes (Wallace et al. 2014). In both cases, the authors reported that while SNPs were most often associated with GWAS hits by virtue of their prevalence, CNVs are the most highly enriched polymorphism class in GWAS hits relative to their genome-wide frequency. In another study that conducted GWAS for key developmental transitions including the juvenile-to-adult vegetative and the vegetative-to-reproductive transitions, it was shown that novel gene associations were identified using transcript abundance and transcript PAV as markers relative to analyses that used only SNP markers (Hirsch et al. 2014). Presumably, some of the transcript PAV markers used in this study are based on genomic level PAV. A comparison of two maize *de novo* genome assemblies and the transcriptome profiles across six tissues from these genotypes revealed that approximately half of the transcript PAVs that were observed were the product of genome-level PAV (Hirsch et al. 2016). Furthermore, a broad-scale study across more than 14,000 maize inbred lines found that phenotypic variation in four complex traits was more associated with SNPs linked to PAVs than to SNPs not linked to PAVs (Lu et al. 2015). Finally, a diversity characterization of maize landraces found that the majority of SNPs associated with altitude adaptation overlapped regions of the genome with large-scale structural variation (Romero Navarro et al. 2017).

Despite the extensive levels of PAV and CNV detected across maize and the enrichment of structural variation in GWAS hits, there are relatively few examples of well-characterized phenotypes in maize that result from a specific structural variant (Table 2.1). One of the first examples of a structural variant affecting a phenotype in maize was enhanced aluminum tolerance resulting from copy number amplification of the MATE1 gene, a transporter from the multidrug and toxic compound extrusion family (Maron et al. 2013). The tunicate phenotype of pod corn (*Zea mays* var *tunicate*) is another example of a structural variant affecting a developmental phenotype (Wingen et al. 2012; Han et al. 2012). The characteristic phenotype of glume covered kernels in the *Tunicate1* (*Tu1*) mutant is the result of ectopic expression of *Zmm19*, a MADS box transcription factor, expressed in developing maize inflorescence. The ectopic expression of *Zmm19* is manifested through a ~1.8 MB inversion associated with a *Mutator-like* transposon. A more extreme tunicate phenotype caused by duplication of two

genes at the breakpoint of the rearrangement can also be seen. The *White Cap* (*Wc*), locus in maize, is another example of structural variation brought about through transposon rearrangement (Tan et al. 2017). Variable repeats of a carotenoid-degrading enzyme, *Ccd1*, at this locus confer quantitative variation for grain color and are the basis for the white-endosperm phenotype. Another example of a structural variant associated with a mutant carotenoid phenotype is the Maize *white seedling 3* (*w3*) locus. This classical mutant phenotype was recently shown to be caused by a complete gene deletion of a homogentisate solanesyl transferase (HST) gene (Hunter et al. 2018). Finally, at the *sugary enhancer* (*Se1*) gene that is important for fresh market sweet corn, there is a recessive allele (*se1*) that is a 630 bp deletion, which eliminates the entire open reading frame of *Se1* and results in loss of normal *Se1* transcript and function. The recessive allele in combination with *sugary1* results in increased sugar content and high levels of water-soluble polysaccharide in the endosperm (Haro von Mogel et al. 2013).

While there are only a few examples of cloned genes in maize with natural PAV/CNV alleles, there are numerous other examples across the plant kingdom (Table 2.1). These cloned examples in other species have a range of phenotypic outcomes from biotic/abiotic stress tolerance to developmental impacts and production of novel secondary metabolites. The technological advances described earlier are decreasing the barriers to *de novo* genome assembly, which will facilitate CNV and PAV discovery and reduce the recalcitrant nature of studying the phenotypic outcomes of these genomic features. It is anticipated that as multiple reference genome assemblies become available for various plant species, including maize, the ability to identify and characterize functional structural variants will improve.

2.5.3 Heterosis

Since the discovery of interspecific gene content variation in maize, there has been considerable interest in the potential role of variable genes in heterosis. Here, we define heterosis in the breeding context as the difference in performance of a hybrid relative to the performance of its better inbred parent, otherwise known as better parent heterosis. Many non-mutually exclusive hypotheses have been put forward to explain the mechanism of heterosis in maize (reviewed in Kaeppler 2012; Schnable and Springer 2013). The three classical quantitative genetics hypotheses include dominance, overdominance, and epistasis. The dominance hypothesis, which posits that heterosis results from the complementation of mildly deleterious alleles present in inbred parents, is most often invoked in the context genome content variation.

Based on early Sanger sequencing work, it was hypothesized that maize genotypes with complementary dispensable gene subsets would produce hybrid offspring with a more complete suite of quantitative-effect dispensable genes (Fu and Dooner 2002). One of the reasons for invoking gene content variation in discussions of heterosis is that it is consistent with the breeding practice of crossing inbreds from complementary heterotic groups to form superior hybrids. Crosses between opposite pools (i.e., Stiff Stalk Synthetic x Non-Stiff Stalk Synthetic) would be expected to generate a more full gene complement compared to crosses that take place within heterotic group crosses (i.e., Stiff Stalk Synthetic x Stiff Stalk Synthetic). This model was supported by later work that demonstrated patterns in PAVs that reflect heterotic groups (Lai et al. 2010; Hansey et al. 2012). Lai et al. resequenced six elite Chinese breeding lines and found that many of the structural variants identified were private to a single heterotic group (Lai et al. 2010). A second study, based on RNA-seq of 21 diverse North American breeding lines, found 145 loci absent from B73 that also showed heterotic group patterning (Hansey et al. 2012). Further, in a comparison of two *de novo* assemblies from genotypes that have high specific combining ability, over 2,500 PAVs were identified as well as extreme expansion and contraction of gene families (Hirsch et al. 2016). While this association is suggestive, clear evidence for a causal role of gene content variation in heterosis has yet to be realized.

Table 2.1 Examples of copy number variants (CNVs) and presence/absence variants (PAVs) with known phenotypic outcomes

Species	Variant type	Trait	Reference
Barley	CNV	Boron toxicity tolerance	(Sutton et al. 2007)
Barley	CNV	Freezing tolerance	(Knox et al. 2010)
Barley	CNV	Flowering time	(Nitcher et al. 2013)
Cucumber	CNV	Reproductive morphology	(Zhang et al. 2015)
Maize	CNV	Tunicate phenotype	(Wingen et al. 2012; Han et al. 2012)
Maize	CNV	Aluminum tolerance	(Maron et al. 2013)
Maize	CNV	Grain color	(Tan et al. 2017)
Maize	PAV	Carotenoid synthesis	(Hunter et al. 2018)
Opium poppy	PAV	Noscapine synthesis	(Winzer et al. 2012)
Palmer amaranth	CNV	Glyphosate resistance	(Gaines et al. 2010)
Rice	PAV	Phosphorus uptake	(Schatz et al. 2014)
Rice	PAV	Submergence tolerance	(Schatz et al. 2014)
Soybean	CNV	SCN resistance	(Cook et al. 2012)
Tomato	CNV	Fruit size	(Xiao et al. 2008)
Wheat	CNV	Photoperiod response	(Díaz et al. 2012)
Wheat	CNV	Dwarfing	(Li et al. 2012)
Wheat	CNV	Freezing tolerance	(Zhu et al. 2014)
Wheat	CNV	Winter hardiness	(Würschum et al. 2017)

2.5.4 Dosage Balance

The concept of dosage balance has been formalized as the Gene Balance Hypothesis, which declares that balanced stoichiometry among members of multi-subunit complexes is critical for optimal function of the macromolecular complex (Birchler and Veitia 2007, 2010, 2012). In practical terms, this posits that gene products that function as part of a complex or interact closely within a certain biochemical framework will likely have an optimal ratio of subunits. Any change that modifies this ratio, such as alteration of gene copy number, will cause a deviation from the optimal balance. This can have important implications for gene expression regulation and, in the context of this chapter, on the evolutionary fate of CNVs. One line of evidence supporting this hypothesis comes from the study of genes retained in duplicate following the most recent polyploidization event in maize and other paleopolyploids. It has been shown that functional classes of genes that participate in macromolecular complexes such as transcription factors and signaling components are more likely to be retained than other functional classes (Woodhouse et al. 2010). This bias also extends to non-polyploidy-derived copy number polymorphisms. Given that many CNVs segregate, inbreds may contain a more dramatic shift from optimal dosage when averaged across the genome compared to the hybrid state due to complementation. Under this model of heterosis, increased inbred performance is expected to lead to decline in the number CNVs observed across the genome (Kaeppler 2012).

2.6 Future Bioinformatic Challenges in the Era of Multiple Genome Assemblies

The number of sequenced and assembled plant genomes is growing at an exponential rate (Michael and Jackson 2013), and many species,

including maize, have genome assemblies from multiple individuals within the species. This burst of activity is due to the realization that a single reference genome is not representative of the variation present in a species. The availability of additional reference genomes will greatly facilitate structural variation characterization and lead to a better understanding of the maize pan-genome. Before new genomic resources can be effectively used, however, current bioinformatic workflows need to be modified to accommodate multiple reference genomes. Some questions raised by the Computational Pan-Genomics Consortium (Computational Pan-Genomics Consortium 2016) include:

1. *What is a reference genome? The genome of a selected individual, the consensus sequence from a population,* or *a maximal genome with all sequences detected?*
2. *How do we efficiently translate coordinates and compare genome features from one genome assembly to another genome assembly*?
3. *Should we abandon the concept of single, linear reference genome and move toward a graph-based approach?*

The incorporation of alternative/novel loci is an important step toward more comprehensive representation of sequence diversity. One challenge associated with their adoption is that read mapping software must be modified to support alternate loci. The development of "alt-aware" algorithms is an area of extensive development. While these loci are useful for capturing variation at regions of interest, they do not attempt to fully represent variation at the pan-genome level. In order to best utilize the full suite of variation present in a population, research communities will need to move beyond the representation of reference genomes as linear strings. The idea of adopting a graph-based genome has been advanced by the Computational Pan-Genomics Consortium, which has advocated for a paradigm shift in how we think of reference genomes. Graph-based structures are already commonly used in assembly software in the form of de Bruijn graphs, which are directed graph structures in which nodes represent *kmers* (unique strings of length *k*) and edges represent an overlap of *k-1* bases between two nodes. Similarly, a basic graph structure might encode shared sequences as nodes in a graph and novel sequences as edges.

Moving from a reference genome being a linear representation of single genotype to a graph-based data structure that represents an amalgam of haplotypes will require new a consensus data structure, new coordinate systems, and the modification of genome browsers and other tools. However, as additional genome assemblies become available and our knowledge of the size and complexity of species pan-genomes continues to grow, the difficulty in these challenges will be far outweighed by the benefit to biological understanding and utilization of diversity in plant species.

Acknowledgements This work was funded in part by the National Science Foundation (Grant IOS-1546727) and ABB was supported by the DuPont Pioneer Bill Kuhn Honorary Fellowship and the University of Minnesota MnDRIVE Global Food Ventures Graduate Fellowship.

References

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90:7980–7984

Albert PS, Gao Z, Danilova TV, Birchler JA (2010) Diversity of chromosomal karyotypes in maize and its relatives. Cytogenet Genome Res 129:6–16

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. Nature 12:363–376

Anderson JE, Kantar MB, Kono TY, et al (2014) A roadmap for functional structural variants in the soybean genome. G3-Genes Genom Genet 4:1307–1318

Ayonoadu UW, Rees H (1971) Effects of B chromosomes on the nuclear phenotype in root meristems of maize. Heredity 27:365–383

Bejarano ER, Khashoggi A, Witty M, Lichtenstein C (1996) Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proc Natl Acad Sci USA 93:759–764

Beló A, Beatty MK, Hondred D et al (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120:355–367

Bennetzen JL, Ramakrishna W (2002) Exceptional haplotype variation in maize. Proc Natl Acad Sci USA 99:9093–9095

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530

Berglund J, Nevalainen EM, Molin A-M et al (2012) Novel origins of copy number variation in the dog genome. Genome Biol 13:R73

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19:395–402

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol 186:54–62

Birchler JA, Veitia RA (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci USA 109:14746–14753

Brohammer AB, Kono TJY, Springer NM et al (2018) The limited role of differential fractionation in genome content variation and function in maize (*Zea mays* L.) inbred lines. Plant J 93:131–141

Brown WL (1949) Numbers and distribution of chromosome knobs in united states maize. Genetics 34:524–536

Brunner S, Fengler K, Morgante M et al (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17:343–360

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Curr Opin Plant Biol 9:172–176

Buescher PJ, Phillips RL, Brambl R (1984) Ribosomal RNA contents of maize genotypes with different ribosomal RNA gene numbers. Biochem Genet 22:923–930

Burr B, Burr FA, Matz EC, Romero-Severson J (1992) Pinning down loose ends: mapping telomeres and factors affecting their length. Plant Cell 4:953–960

Cao J, Schneeberger K, Ossowski S et al (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43:956–963

Chia J-M, Song C, Bradbury PJ et al (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807

Computational Pan-Genomics Consortium (2016) Computational pan-genomics: status, promises and challenges. Brief Bioinformatics bbw089

Cook DE, Lee TG, Guo X et al (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science 338:1206–1209

Darracq A, Vitte C, Nicolas S et al (2018) Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. BMC Genom 19:119

Dietrich CR, Perera MADN, D Yandeau-Nelson M, et al (2005) Characterization of two GL8 paralogs reveals that the 3-ketoacyl reductase component of fatty acid elongase is essential for maize (Zea mays L.) development. Plant J 42:844–861

Díaz A, Zikhali M, Turner AS et al (2012) Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). PLoS One 7:e33234

Emrich SJ, Li L, Wen T-J et al (2007) Nearly identical paralogs: implications for maize (Zea mays L.) genome evolution. Genetics 175:429–439

Flavell RB (1986) Repetitive DNA and chromosome evolution in plants. Philos Trans R Soc London/Ser B 312:227–242

Force A, Lynch M, Pickett FB et al (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Freeman VJ (1951) Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae. J Bacteriol 61:675–688

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gaines TA, Zhang W, Wang D et al (2010) Gene amplification confers glyphosate resistance in amaranthus palmeri. Proc Natl Acad Sci USA 107:1029–1034

Genomes Consortium (2016) 1,135 Genomes reveal the global pattern of polymorphism in arabidopsis thaliana. Cell 166:481–491

Golicz AA, Bayer PE, Barker GC et al (2016) The pangenome of an agronomically important crop plant Brassica oleracea. Nat Commun 7:13390

Gore MA, Chia J-M, Elshire RJ et al (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin AC, Schackwitz W, Tyler L, Martin J (2017) Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. Nature Commun 8:2184

Hake S, Walbot V (1980) The genome of Zea mays, its organization and homology to related grasses. Chromosoma 79:251–270

Han J-J, Jackson D, Martienssen R (2012) Pod corn is caused by rearrangement at the Tunicate1 locus. Plant Cell 24:2733–2744

Hansey CN, Vaillancourt B, Sekhon RS et al (2012) Maize (Zea mays L.) genome diversity as revealed by RNA-sequencing. PLoS One 7:e33071

Haro von Mogel K, Hirsch CN, De Vries B et al (2013) The mapping, genetic analysis, and phenotypic

characterization of sugary enhancer1 (se1). Maize Genet Conf Abs 55:T16
Hirsch CN, Foerster JM, Johnson JM et al (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26:121–135
Hirsch CN, Hirsch CD, Brohammer AB et al (2016) Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. Plant Cell 28:2700–2714
Hogg JS, Hu FZ, Janto B et al (2007) Characterization and modeling of the haemophilus influenzae core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol 8:R103
Hunter CT, Saunders JW, Magallanes-Lundback M, Christensen SA, Willett D, Stinard PS, Li Q-B, Lee K, DellaPenna D, Koch KE (2018) Maize w3 disrupts homogentisate solanesyl transferase (ZmHst) and reveals a plastoquinone-9 independent path for phytoene desaturation and tocopherol accumulation in kernels. Plant J 93:799–813
Jacq C, Miller JR, Brownlee GG (1977) A pseudogene structure in 5S DNA of xenopus laevis. Cell 12:109–120
Jeffares DC, Jolly C, Hoti M et al (2017) Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun 8:14061
Jiao Y, Peluso P, Shi J, et al (2017) Improved maize reference genome with single molecule technologies. bioRxiv
Kaeppler SM (2012) Heterosis: many genes, many mechanisms—end the search for an undiscovered unifying theory. ISRN Bot. 2012:682824
Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci USA 101:13554–13559
Kato TAY (1976) Cytological studies of maize [Zea mays L.] and teosinte [Zea mexicana Schrader Kuntze] in relation to their origin and evolution. Univ. Mas. Agric. Expt. Sta
Knox AK, Dhillon T, Cheng H et al (2010) CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. Theor Appl Genet 121:21–35
Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36:6688–6719
Kyndt T, Quispe D, Zhai H et al (2015) The genome of cultivated sweet potato contains Agrobacterium T-DNAs with expressed genes: an example of a naturally transgenic food crop. Proc Natl Acad Sci USA 112:5844–5849
Lai J, Li R, Xu X et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:1027–1030
Lai J, Li Y, Messing J, Dooner HK (2005) Gene movement by helitron transposons contributes to the haplotype variability of maize. Proc Natl Acad Sci USA 102:9068–9073
Langham RJ, Walsh J, Dunn M et al (2004) Genomic duplication, fractionation and the origin of regulatory novelty. Genetics 166:935–945
Laurie DA, Bennett MD (1985) Nuclear DNA content in the genera zea and sorghum. intergeneric, interspecific and intraspecific variation. Heredity 55:307–313
Li R, Li Y, Zheng H et al (2010) Building the sequence map of the human pan-genome. Nat Biotechnol 28:57–63
Li Y, Xiao J, Wu J et al (2012) A tandem segmental duplication (TSD) in green revolution gene Rht-D1b region underlies plant height variation. New Phytol 196:282–291
Li Y-H, Zhou G, Ma J et al (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol 32:1045–1052
Liu S, Ying K, Yeh C-T et al (2012) Changes in genome content generated via segregation of non-allelic homologs. Plant J 72:390–399
Liu S, Zheng J, Migeon P et al (2017) Unbiased K-mer analysis reveals changes in copy number of highly repetitive sequences during maize domestication and improvement. Sci Rep 7:42444
Lu F, Romay MC, Glaubitz JC et al (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. Nat Commun 6:6914
Maize Genetics and Genomics Database. Information about maize assembly Zm-Mo17-REFERENCE-NRGENE-1.0. In: maizegdb.org. http://maizegdb.org/genome/genome_assembly/Zm-Mo17-REFERENCE-NRGENE-1.0. Accessed 2 Jun 2017
Maize Genetics and Genomics Database. Information about maize assembly Zm-B104-DRAFT-ISU_USDA-0.1. In: maizegdb.org. https://maizegdb.org/genome/genome_assembly/Zm-B104-DRAFT-ISU_USDA-0.1. Accessed 2 Jun 2017
Maize Genetics and Genomics Database. Maize CML247. In: maizegdb.org. http://maizegdb.org/gbrowse/maize_cml247. Accessed 2 Jun 2017
Makarevitch I, Waters AJ, West PT et al (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet 11:e1004915
Maron LG, Guimarães CT, Kirst M et al (2013) Aluminum tolerance in maize is associated with higher MATE1 gene copy number. Proc Natl Acad Sci USA 110:5241–5246
McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36:344–355
McClintock B, Kato A, Blumenschein A (1981) Chromosome constitution of races of maize: its significance

in the interpretation of relationship between races and varieties in the Americas. Mexico: Colegio de Postgraduados, p 517
Medini D, Donati C, Tettelin H et al (2005) The microbial pan-genome. Curr Opin Genet Dev 15:589–594
Messing J, Bharti AK, Karlowski WM et al (2004) Sequence composition and genome organization of maize. Proc Natl Acad Sci USA 101:14349–14354
Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 11:1660–1676
Michael TP, Jackson S (2013) The first 50 plant genomes. Plant Genome 6:2
Montenegro JD, Golicz AA, Bayer PE et al (2017) The pangenome of hexaploid bread wheat. Plant J 90:1007–1013
Morgante M, Brunner S, Pea G et al (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002
Näsvall J, Sun L, Roth JR, Andersson DI (2012) Real-time evolution of new genes by innovation, amplification, and divergence. Science 338:384–387
Nitcher R, Distelfeld A, Tan C et al (2013) Increased copy number at the HvFT1 locus is associated with accelerated flowering time in barley. Mol Genet Genomics 288:261–275
Ohno S (1970) Evolution by gene duplication. Springer-Verlag, New York, USA
Peacock WJ, Dennis ES (1981) Highly repeated DNA sequence limited to knob heterochromatin in maize. Proc Natl Acad Sci USA 78:4490–4494
Phillips RL, Weber DF, Kleese RA, Wang SS (1974) The Nucleolus organizer region of maize (Zea mays L.): tests for ribosomal gene compensation or magnification. Genetics 77:285–297
Poggio L, Rosato M, Chiavarino AM, Naranjo CA (1998) Genome size and environmental correlations in maize (Zea mays ssp. mays, poaceae). Ann Bot 82:107–115
Rayburn AL, Price HJ, Smith JD, Gold JR (1985) C-Band Heterochromatin and DNA content in zea mays. Am J Bot 72:1610–1617
Rivin CJ, Cullis CA, Walbot V (1986) Evaluating quantitative variation in the genome of zea mays. Genetics 113:1009–1019
Romero Navarro JA, Wilcox M, Burgueño J et al (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet 49:476–480
SanMiguel PJ, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 82:37–44
Schatz MC, Maron LG, Stein JC et al (2014) Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. Genome Biol 15:506
Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA 108:4069–4074
Schnable PS, Springer NM (2013) Progress toward understanding heterosis in crop plants. Annu Rev Plant Biol 64:71–88
Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115
Song R, Messing J (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci USA 100:9055–9060
Springer NM, Ying K, Fu Y et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5:e1000734
Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet 43:1160–1163
Sutton T, Baumann U, Hayes J et al (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. Science 318:1446–1449
Swanson-Wagner RA, Eichten SR, Kumari S et al (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20:1689–1699
Syvanen M (2012) Evolutionary implications of horizontal gene transfer. Annu Rev Genet 46:341–358
Tan B-C, Guan J-C, Ding S et al (2017) Structure and Origin of the White Cap Locus and its role in evolution of grain color in maize. Genetics 206:135–150
Tang H, Bowers JE, Wang X et al (2008) Synteny and collinearity in plant genomes. Science 320:486–488
Tenaillon MI, Hollister JD, Gaut BS (2010) A triptych of the evolution of plant transposable elements. Trends Plant Sci 15:471–478
Tettelin H, Masignani V, Cieslewicz MJ et al (2005) Genome analysis of multiple pathogenic isolates of streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci USA 102:13950–13955
Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11:472–477
Tikhonov AP, SanMiguel PJ, Nakajima Y et al (1999) Colinearity and its exceptions in orthologous adh regions of maize and sorghum. Proc Natl Acad Sci USA 96:7409–7414
Torres EM, Williams BR, Amon A (2008) Aneuploidy: cells losing their balance. Genetics 179:737–746
Unterseer S, Seidel MA, Bauer E, Haberer G (2017) European flint reference sequences complement the maize pan-genome. bioRxiv
Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. Nat Rev Genet 408:796
Vielle-Calzada J-P, Martínez de la Vega O, Hernández--Guzmán G, et al (2009) The Palomero genome

suggests metal effects on domestication. Science 326:1078–1078

Wallace JG, Bradbury PJ, Zhang N et al (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet 10: e1004845

Wetterstrand KA (2018) DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP). www.genome.gov/sequencingcostsdata. Accessed 17 Apr

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130

Wingen LU, Münster T, Faigl W et al (2012) Molecular genetic basis of pod corn (Tunicate maize). Proc Natl Acad Sci USA 109:7115–7120

Winzer T, Gazda V, He Z et al (2012) A papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science 336:1704–1708

Woodhouse MR, Schnable JC, Pedersen BS et al (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. PLoS Biol 8:e1000409–e1000415

Würschum T, Longin CFH, Hahn V et al (2017) Copy number variations of CBF genes at the Fr-A2 locus are essential components of winter hardiness in wheat. Plant J 89:764–773

Xiao H, Jiang N, Schaffner E et al (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319:1527–1530

Yandeau-Nelson MD, Xia Y, Li J et al (2006) Unequal sister chromatid and homolog recombination at a tandem duplication of the A1 locus in maize. Genetics 173:2211–2226

Yandeau-Nelson MD, Zhou Q, Yao H et al (2005) MuDR transposase increases the frequency of meiotic crossovers in the vicinity of a mu insertion in the maize a1 gene. Genetics 169:917–929

Yao W, Li G, Zhao H et al (2015) Exploring the rice dispensable genome using a metagenome-like assembly strategy. Genome Biol 16:187

Yona AH, Manor YS, Herbst RH et al (2012) Chromosomal duplication is a transient evolutionary solution to stress. Proc Natl Acad Sci USA 109:21010–21015

Yoshida S, Maruyama S, Nozaki H, Shirasu K (2010) Horizontal gene transfer by the parasitic plant Striga hermonthica. Science 328:1128–1128

Young ND, Zhou P, Silverstein KA (2016) Exploring structural variants in environmentally sensitive gene families. Curr Opin Plant Biol 30:19–24

Zhang Z, Mao L, Chen H et al (2015) Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. Plant Cell 27:1595–1604

Zhou P, Silverstein KAT, Ramaraj T et al (2017) Exploring structural variation and gene family architecture with de novo assemblies of 15 medicago genomes. BMC Genom 18:261

Zhu J, Pearce S, Burke A et al (2014) Copy number and haplotype variation at the VRN-A1 and central FR-A2 loci are associated with frost tolerance in hexaploid wheat. Theor Appl Genet 127:1183–1197

3 Rapid, Affordable, and Scalable Genotyping for Germplasm Exploration in Maize

M. Cinta Romay

Abstract

Over the last few decades, next-generation sequencing has revolutionized the way that genetic diversity is preserved, studied, and used. Maize, because it is one of the most important crops in the world, has been at the cutting edge in the application of these new technologies. Recently, several different platforms have been used to explore the diversity of the maize genome and its connection to observed phenotypes. This chapter presents a summary of the most successful platforms to date. As technology improves at a rapid rate, generating new and more complex data, researchers need new tools to optimize the use of resources to explore maize germplasm for genomics and breeding.

3.1 Introduction

Maize (*Zea mays L.)* is one of the major crops in the world. It is cultivated in most of the temperate and tropical regions of the planet and is a main source for human food, animal feed, and industrial processes (FAOSTAT 2018). Moreover, it is an important model plant for the scientific community used to study many different biological processes. The economic and scientific importance of maize makes it one of the first crops where cutting-edge technologies are usually applied. From the first studies using isoenzymes to characterize maize diversity (Stuber and Goodman 1983) or analyzing quantitative traits (Edwards et al. 1987), to the newest next-generation technologies for an improved reference genome (Jiao et al. 2017a), maize has a rich history of developing community resources generated using innovative molecular techniques.

Molecular markers in maize are used for population and genome evolution studies, germplasm characterization and selection of core collections, trait mapping, and breeding. For example, in a market where intellectual property rights for breeders need to be protected, essentially derived varieties (EDV) are identified using a set of 435 single sequence repeat (SSR) markers (Kahler et al. 2010). This allows breeders to determine if someone has stolen their variety. Similarly, genebanks can use molecular markers to analyze their collections and help make decisions regarding contamination, errors, or duplicated entries while they minimize the negative effects of genetic drift or unconscious selection, and maximize use of resources (Romay et al. 2013). If the information generated by the banks is shared, breeders can use it in combination with passport and phenotypic data to identify sources

M. C. Romay (✉)
Institute for Genomic Diversity, Cornell University, 175 Biotechnology Bldg., Ithaca 14853, USA
e-mail: mcr72@cornell.edu

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_3

of possible favorable alleles to introduce into their programs through biotechnology or pre-breeding (Wang et al. 2017). The same information can also be used to study the relationship between genotypes and phenotypes and better understand the genetic architecture of quantitative traits (Romero Navarro et al. 2017). Furthermore, markers scored on ancient DNA combined with information on recent accessions can help us predict ancient phenotypes and understand the dynamics of adaptation (Swarts et al. 2017). In addition, the prediction of phenotypes from only genotypic information using genomic selection (GS) has revolutionized modern breeding (Meuwissen et al. 2001). Maize breeding programs already benefit from the use of molecular markers in GS (Xu et al. 2017b; Han et al. 2018) and even greater progress can be made in the public sector if we achieve low-cost and ultra high-throughput genotyping (Buckler et al. 2016; Rasheed et al. 2017).

Over the last three decades, the maize community has been using different types of DNA-based markers for their work. Starting with

Table 3.1 Commonly used genotyping methods in maize

Method	Advantages	Disadvantages
KASP™ and TaqMan™	• Low cost • Work in low-quality DNA • Quick turnaround time	• Very low throughput (<10 SNPs/sample)
Illumina® GoldenGate™ and MaizeSNP50 BeadChip	• Stable and defined set of SNPs • Very low missing • Reliable call of hets	• Middle throughput (thousands SNPs/sample) • Expensive • Ascertainment bias
Affymetrix® Axiom® Maize genotyping arrays	• Stable and defined set of SNPs • Very low missing data • Reliable call of hets • High throughput (hundreds of thousands SNPs/sample)	• Expensive • Ascertainment bias
Genotyping by sequencing (GBS)	• Reduced cost per sample • Less biased (no discovery step needed) • High throughput (hundreds of thousands SNPs/sample)	• High missing data • High errors in heterozygote calls • Requires high-quality DNA
DaRTSeq, tGBS, RAPID Seq	• Reduced cost per sample • High throughput (hundreds of thousands SNPs/sample. Less than GBS) • Lower missing data than GBS • Reliable heterozygote calls	• Requires high-quality DNA
rAmpSeq	• Low cost • Work with low-quality DNA • Quick turnaround time • Reliable heterozygote calls	• Middle throughput (thousands markers/sample) • Complicated and species customized bioinformatics • Only dominant markers
Transcriptome sequencing	• Low missing data • Provides additional information (expression)	• Labor intensive • Limited representation of genome • Dependent on tissue sampled • Complicated bioinformatics and data management
Whole genome resequencing	• Very high throughput (millions SNPs/sample) • Less biased (no discovery step needed)	• Complicated bioinformatics and data management • Missing data and errors correlated with cost

SSR and Amplified Fragment Length Polymorphisms (AFLP), the recent development of new sequencing technologies has made single-nucleotide polymorphisms (SNPs) the markers of choice. This is because of their abundance and even distribution across the genome (Gupta et al. 2008; Kumar et al. 2012). An SNP is a single-nucleotide difference in an otherwise identical sequence of the genome. The use of SNP markers requires an initial discovery step where a set of diverse accessions is sequenced and compared. Once discovered, SNPs can be scaled to use in many different methods (Gupta et al. 2008). However, the maize genome is large and complex when compared with other crops and model plant species (Fu and Dooner 2002). The levels of diversity between two corn inbred lines are higher than between humans and chimpanzees (Buckler et al. 2006). This diversity is caused by a dynamic genome (Hirsch et al. 2014b) and a number of structural variations that often associate with phenotypic changes (Wallace et al. 2014). Some of these structural variations include location of repetitive elements and genes (Morgante et al. 2005; Chia et al. 2012), presence/absence variation (PAV), and copy number variations (CNV) (Springer et al. 2009), and even large differences in chromosomal structure like big inversions that are related to phenotypic variation and adaptation (Pyhäjärvi et al. 2013; Romero Navarro et al. 2017). These features make the development of SNP-based genotyping platforms for maize more challenging than for humans, animals, and other crop species with smaller genomes.

Regardless, there are several successful platforms generally used for genotyping in maize. Some of the most popular methods are single SNP methods like KASP™ from LGC Genomics (He et al. 2014) and TaqMan™ from Applied Biosystems (Tobler et al. 2005). However, there are also other systems that can produce a few thousand to a hundred thousand SNPs per sample, like array-based technologies such as BeadXpress™ and GoldenGate™ from Illumina® (Yan et al. 2010; Ganal et al. 2011) and Axiom® Genotyping Array from Affymetrixs® (Unterseer et al. 2014) or next-generation sequence-based genotyping platforms, such as DArTSeq, developed by Diversity Arrays Technology (Sansaloni et al. 2011), or genotyping by sequencing (GBS) (Elshire et al. 2011). These approaches can be used for different applications, depending mostly on the amount of data needed, available cost per sample, and tolerance of missing data. This chapter gives an overview of the resources available to the community to characterize maize genetic diversity in a high-throughput manner, the uses to date, and future directions. A summary of frequently used methods, their advantages, and disadvantages is presented in Table 3.1.

3.2 Array-Based Genotyping Platforms

Array technologies are designed to target mostly the genic portions of the genome, which may imply targeting more of the SNPs that influence the phenotypes. They are more stable than other technologies, since they always characterize the same set of SNPs across materials and have very little missing data points making comparison and replication of results easier. In return, they are more expensive than other technologies, with a correlation between the price of the array and the number of SNPs it contains, and they often miss most of the regulatory regions. In addition, the need to select only certain sections of the genome for its development requires an initial discovery work and SNP selection that frequently causes reference bias (Ganal et al. 2012).

Most of the genotyping array platforms have been developed based on technologies from Illumina® and Affymetrix®. Although the purpose and output of the different technologies are similar, the process to obtain the data differs significantly (Rasheed et al. 2017). There are several reviews with technological comparisons of these different platforms and their applications in different crops (Gupta 2008). Here, we will focus on the characteristics and use in maize.

3.2.1 Illumina® GoldenGate™ Assay

The GoldenGate assay is a medium density array that incorporates locus and allele-specific oligos that fit into patterned microwells for hybridization, followed by allele-specific extension and fluorescent scanning of 48–384 and 384–3072 SNPs per sample (Shen et al. 2005). In maize, several 1,536 Oligo Pool Assays (OPA) have been developed from resequencing data from the Maize Sequence Project (www.panzea.org) and used for different purposes. The first integrated linkage map of the Nested Association Mapping (NAM) Population with 1,106 SNPs was established with this technology between 2006 and 2008 (McMullen et al. 2009). The mappable SNPs were combined with an additional 430 obtained from the same database and used to test two populations of Recombinant Inbred Lines (RILs) and a diverse collection of 154 inbred lines, most of them from Chinese origins. Approximately 92% of the SNPs tested were called successfully and the genotyping was highly repeatable, showing less than 20% missing data in these materials (Yan et al. 2010). Those high-quality SNPs were later used to genotype diverse materials, most of them important for breeding programs around the world (Yang et al. 2010; Pan et al. 2012). Another OPA was also developed, replacing the unmappable SNPs with markers from candidate genes associated with mechanisms and pathways related to drought tolerance. This chip was tested in a global collection of inbred lines from temperate, tropical, and subtropical breeding programs and its information was used to study Linkage Disequilibrium (LD) decay and to select a core collection that maximized diversity (Yan et al. 2009). This technology was also used to characterize breeding pools from other different programs (Nelson et al. 2008, 2016; Jones et al. 2009; Wen et al. 2011a; Semagn et al. 2012), and to study the integrity of accessions preserved in different germplasm banks (Wen et al. 2011b).

3.2.2 Illumina® MaizeSNP50 BeadChip

The MaizeSNP50 BeadChip uses the same BeadArray technology that the GoldenGate assay, but is expanded to higher density using Infinium® assays. These assays are based on a two-color single base extension from a single hybridization probe per SNP marker with allele calls ranging from 3,000 to over 5 million SNPs per sample (Steemers and Gunderson 2007). The initial design of this array started with the selection of 800,000 SNPs across multiple discovery projects. From those, a set of approximately 58,000 were initially chosen, giving priority to SNPs located in genes. Among the 49,585 scorable SNPs, 34,182 came from Panzea, 13,037 from Syngenta, 1,816 from INRA, 400 from TraitGenetics, and 150 from other sources. These markers covered a total of 17,520 different genes. In general, the markers are well distributed across the genome, although with lower numbers toward the centromeric regions. Usually, the distance between markers is a few kilobases (kb), although there are some gaps up to 100 kb, and even a 2 Megabase (Mb) region without markers on chromosome 6. Ganal et al. (2011) detailed information for the array and each SNP, including source and genotypes across all the initially scored 274 lines. The biases in this array were investigated and compared with other marker systems. The Syngenta SNPs, selected mostly from two inbreds, B73 and Mo17, can be problematic if the array is used for comparisons between genotypes belonging to different heterotic groups. Selection of only the most diverse subset of the markers is recommended in those cases (Ganal et al. 2011; Romay et al. 2013). The MaizeSNP50 BeadChip chip has been widely used for different purposes and materials. For example, it has been used to characterize the MAGIC maize population (Dell'Acqua et al. 2015), to study the genetic architecture of kernel composition (Cook et al. 2012; Li et al. 2013), to characterize two new European association panels

(Rincent et al. 2014), and to study population genomics of adaptation, domestication, and improvement (Hufford et al. 2012a; van Heerwaarden et al. 2012; Pyhäjärvi et al. 2013).

Several smaller and germplasm-specific arrays were developed later by selecting subsets of these SNPs. The GoldenGate array chip maizeSNP3072, containing 3,072 SNPs, was developed to fingerprint and perform marker-assisted breeding in Chinese maize varieties (Tian et al. 2015), and it has been used in several studies involving mostly Chinese germplasm (Li et al. 2014; Yin et al. 2014; Guo et al. 2015; Cui et al. 2015; Hao et al. 2015a, b; Zhou et al. 2016, 2017; Meng et al. 2016; Shi et al. 2016; Song et al. 2017; Hu et al. 2017; Luo et al. 2017). KWS SAAT AG (Einbeck, Germany) has a custom Infinium iSelectHD® chip that comprises 9,000 of the SNPs (Pestsova et al. 2016). Similarly, Limagrain has a design which includes a subset of 18,480 SNPs on a Affimetrix® array (Giraud et al. 2017). A collaboration between Illumina, Syngenta, and Pioneer created a standard tool for EDV testing, featuring 3,047 evenly distributed SNPs (MaizeLD BeadChip 3 K). Nucleotide selection for this chip was based on a comparison of pairwise distance data for public and commercially relevant samples of US and European lines and SSR markers previously shown to discern EDV status (Rousselle et al. 2015).

3.2.3 600 K and 55 K Affymetrix® Axiom® Maize Genotyping Array

Recently Affymetrix implemented the Axiom technology based on a two-color, ligation-based assay with 30-mer probes allowing simultaneous genotyping of 384 samples with 50 K SNPs, or 96 samples with 650 K SNPs (Hoffmann et al. 2011). In maize, a high-density genotyping array with 616,201 variants, including the SNPs of the Illumina® MaizeSNP50 BeadChip, has been developed (Unterseer et al. 2014). This array is optimized for European and American temperate maize. It is one of the largest publicly available genotyping arrays in crop species and supports the need for higher marker resolution in a crop with a large genome and rapid LD decay like maize. Starting with 57 million SNPs and small indels discovered by mapping whole genome sequence reads of 30 representative temperate maize lines, 1.2 million variants were selected for further testing on 285 temperate maize samples. The final selection of variants was based on their high quality in assay performance, physical distribution, and concordance with in silico variant calls from sequencing data. The 600 K array has an average density of one variant every 3.4 kb, with a median of one variant per 0.3 kb. Approximately 89% of genes were represented by at least one variant. This assay has already been used to genotype a collection of European landraces (Mayer et al. 2017) and for a Genotype-by-Environment (G × E) study in European hybrids (Millet et al. 2016).

Another platform with 55,229 SNPs has been recently established (Xu et al. 2017a) with this technology. This array shows lower rates of missing data and heterozygosity and more SNPs with lower minor allele frequencies in tropical materials than the MaizeSNP50 BeadChip. Initial SNPs were selected for even distribution from the 600,000. Then, supplemental SNPs that were highly polymorphic between temperate and tropical materials were added by selecting from the MaizeSNP50 BeadChip and a RNA-Seq study (Fu et al. 2013). SNPs from the dispensable parts of the genome, the classic maize gene list, and tags for published transgenic events completed the selection. The array has been evaluated in a diverse set of approximately 500 inbred lines mostly from China, USA, and the tropics (Xu et al. 2017a), and used to map starch paste viscosity in waxy corn (Hao et al. 2017).

3.3 Sequence-Based Genotyping Platforms

Recent advances in DNA sequencing technologies over the last decade have allowed the production of millions of bases per sample in one round. These improvements have made the use of sequencing for genotyping feasible (Egan

et al. 2012). Sequence-based genotyping has several advantages when compared with array technologies. First, the discovery and SNP calling process can be done in one single step. Second, raw sequence reads can be stored and reevaluated with different analysis pipelines as new bioinformatic tools, improved reference genomes, or new datasets are developed. Third, more calls are obtained across the genome for a smaller cost and the ascertainment bias is reduced since there is no need to select in advance which markers will be scored. On the other hand, bioinformatic analysis of the resulting data is, in general, more challenging, and the set of markers scored on each sample is highly variable depending on many different factors (sample prep, machine, coverage, etc.). Additionally, there are difficulties related to insufficient and variable read depth, which leads to problems with missing data.

Although sequencing the entire genome would always be ideal, it is often not feasible. In the case of large and repetitive genomes like maize, costs can quickly become an issue. A reduced representation approach to get better coverage of sequence reads in targeted regions can be a less expensive approach (Hirsch et al. 2014a; Ott et al. 2017). These samples can be obtained in many ways. The most common procedure is the use of a restriction enzyme (RE) targeted to keep less repetitive regions of the genome. There are different protocols to obtain these type of libraries with different levels of genome reduction, from the initial RAD-Seq (Baird et al. 2008), to double digest RAD-Seq (Peterson et al. 2012), 2b-RAD (Wang et al. 2012), GBS (Elshire et al. 2011), tGBS (Ott et al. 2017), Phased GBS (Manching et al. 2017), DArTSeq (Sansaloni et al. 2011), etc. Other options for a reduced representation of the genome are exome capture (Hodges et al. 2007; Fu et al. 2010; Yang et al. 2015), transcriptome sequencing (Wang et al. 2009), or amplicon sequencing (Bybee et al. 2011). Many of these reduced representation methods have been successfully used in maize, a summary of the most frequently approaches and some of their applications are presented below.

3.3.1 Genotyping by Sequencing (GBS)

The most popular genotyping by sequencing protocol applied to maize was first developed at Cornell University (Elshire et al. 2011). The method avoids repetitive regions of the genome using the *Ape*KI enzyme and targets any sequence with low copy, including the important non-coding regulatory regions. The protocol is easy, quick, reproducible, and specific. The procedure eliminates the size selection and a few enzymatic and purification steps and adds the inexpensive barcoding system to a previously developed method (Baird et al. 2008). The result was a genotyping method that produced more SNPs at a much lower cost than the SNP arrays popular at the time. The protocol was initially tested in maize with 276 RILs from the IBM population (Lee et al. 2002), and later expanded to the entire NAM population (McMullen et al. 2009).

In 2013, GBS was used to explore the entire collection of maize inbred lines preserved by the US national germplasm system (Romay et al. 2013). Although researchers have characterized subsets of these inbred lines for different purposes, this was the first time that technology enabled a large-scale genomic characterization of maize resources and haplotypes. This allowed the discovery of rare variants and a deeper study of LD decay in maize. This GBS data has been used to identify accessions that have been misclassified, select best sources for multiplication and distribution, eliminate duplications, select core collections, add or recommend new experimental entries, and assess genetic profile changes over successive regenerations.

The first analysis of the US national germplasm system sequence data produced approximately 700,000 markers distributed across the genome, with reduced representation in the pericentromeric regions. The mean discrepancy rate when SNP calls were compared with the llumina® MaizeSNP50 BeadChip data was about 0.6% for homozygote calls, while average missing data was 35%. Results showed that most of the SNPs in the collection are rare, present in

less than 5% of the accessions, with some alleles unique to certain groups of germplasm. Only a modest amount of the total genetic diversity was present in the pool of Ex-PVP lines. GBS data also allowed the study of the integrity of the USDA maize collection, showing that more than 98% of samples with the same name were the same. GBS permitted the reconstruction of the expected pedigree and breeding program relationships and showed less ascertainment bias than the MaizeSNP50, although still more than previous studies with SSR markers. The bias is caused by the step that requires alignment to the reference genome and some of the filtering steps (Romay et al. 2013). This genetic characterization of the collection is publicly available through MaizeGDB (https://www.maizegdb.org/snpversity).

However, as the number of samples processed with GBS increased, an improved bioinformatics pipeline was needed. The key features of the new pipeline were scalability, running from a few to thousands of samples, ability to run on simple small computers with 8–12 GB of RAM usually available to small breeders, rapid turnover, and accommodation of high genomic diversity that is usually encountered in the most important crop species, especially maize (Glaubitz et al. 2014). The new pipeline includes two different options, one for discovery of SNPs and another for production SNP calling. The pipeline favors calling large numbers of SNPs versus depth per SNP and later uses filters based on population genetics to reduce error rate. In maize, the latest discovery build (GBS 2.7) has been done with 31,978 samples (plus 758 blank negative samples) and it produced almost one million SNPs. All of the USDA collection and NAM data was included in this build and data is available through Panzea (https://www.panzea.org/genotypes). In addition, the same pipeline has been used to call SNPs in the collection of inbred lines that have been evaluated from 2014 to 2017 as part of the Genomes2Fields initiative, and the data is publicly available through their Webpage (https://www.genomes2fields.org/resources).

In 2016, the collection of inbred lines at CIMMYT consisting of 539 inbred lines released between 1984 and 2003 was also characterized using GBS 2.7 (Wu et al. 2016). Due to their wide distribution and great contribution to tropical maize breeding improvement, the collection represents the total genetic diversity of improved tropical maize germplasm. The population structure analysis showed that most of the inbreds could be classified into three main clusters: Lowland Tropical, Subtropical/Mid-altitude, and Highland Tropical, with most of them distantly related to each other. Gene diversity of the three tropical subgroups was similar and higher than temperate materials. LD decay was also faster. Public data for this collection can be obtained on the CIMMYT Research Data Repository Website (https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10423).

Another interesting GBS 2.7 public dataset from CIMMYT consists of a large collection of maize landraces from 35 different countries across the Americas. Data can be accessed from CIMMYT Seeds of Discovery Repository (https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10034). Markers were used to study flowering time and its relationship with the genes driving adaptation to low, middle, and high elevation (Romero Navarro et al. 2017). Country of origin was the main clustering factor for the landraces, but sampling across different Mexican altitudes did not show a complete differentiation of the materials, indicating that alleles still segregate between different adaptation classes at the maize center of origin. Latitude and altitude were used to map adaptation and they found that more than 50% of significant hits were in regions of low recombination, in particular a 13 Mb inversion on chromosome 4 introgressed from highland teosinte (Pyhäjärvi et al. 2013). Outside of these regions, numerous other genes associated with adaptation to altitude were identified and most of them were shared across clades and landraces. This study showed the power and resolution that genotypic characterization of landraces and association with well-curated passport data

can have, and has opened the door to further study the materials preserved at the germplasm banks beyond inbred lines.

A collection of 1,197 flint inbred lines representing European diversity has also been characterized using the same GBS pipeline and its diversity compared against the USDA collection (Gouesnard et al. 2017). The inbred lines were selected from the INRA collection, the Cornfed flint association panel (Rincent et al. 2014), and inbred lines recently derived from European landraces. Seven groups, in agreement with known breeding knowledge, were identified. Consistent with previous results, the homozygote data had high quality when compared with previous data from the MaizeSNP50 BeadChip. The collection was compared to the USDA data and, as expected, it showed reduced diversity, although most of the European flints formed their own group (Gouesnard et al. 2017). Although data is not publicly available, it can be obtained upon request.

The main weaknesses of GBS datasets generated with this method are the amount of missing data and undercalling of heterozygote SNPs, both caused by low and uneven read depth. It has been shown in maize that even at relative low coverage, GBS can produce enough information for powerful QTL mapping in biparental populations. However, dense genotyping requires increased target coverage per individual (Beissinger et al. 2013). When working with inbred lines, the depth problem is partially dealt with by analyzing large numbers of samples that share the same haplotype. Once a certain read depth has been achieved for a sample, a big portion of the missing data most likely has real biological value, since it is probably caused by large divergence of the reads with the reference genome, small insertions or deletions, PAV, or larger structural variants. For example, the popcorn inbred line SA24 was sequenced in 25 pools in the USDA collection but still showed 16% missing data (Romay et al. 2013). Similarly, markers that are scored in many more samples than the average or present higher heterozygote rates will probably present higher read depth than the average, and should be used with caution since they are very likely targeting regions of the genome that include CNV.

If missing data is still a problem, modifications of the method can be used to genotype large samples taken from heterogeneous populations of heterozygote materials (Manching et al. 2017; Ott et al. 2017), or multiple imputation procedures can be tested for accuracy. Most of the available imputation pipelines like Beagle v4.0 (Browning and Browning 2016) have been optimized for humans. Although this may work well with highly heterozygous materials, these algorithms cannot take advantage of the library of high-quality scored haplotypes created by inbred lines, large LD blocks present in most crop species, or the ability to use controlled mating designs. In addition, they usually fill all missing data with a variant, losing some of the biological signal present in the initial dataset. In crops with controlled mapping designs, highly accurate recombination points can be found using algorithms like FSFHap for ordered genotypes, or FILLIN, a more generalized method (Swarts et al. 2014). To avoid problems caused by pedigree tracking errors and contamination, both methods do not require known parental genotypes and rely instead on a Hidden Markov Model. Ordered markers can be modeled as a Markov chain and the missing genotype calculated using a Vitervi algorithm (Rabiner 1989). Compared with Beagle v4.0, FILLIN is faster and makes less errors imputing the minor allele on inbreed samples. For highly heterozygous samples like landraces, Beagle v4.0 performs better. Regardless of the materials or imputation method, the generation of a subset of genotypes covering the diversity of haplotypes at a higher coverage to ensure good haplotype generation is recommended (Swarts et al. 2014).

3.3.2 DArTseq

Another protocol that uses restriction enzymes for complexity reduction is DArTseq (Diversity Arrays Technology Pty. Ltd. 2018). While the initial method involved hybridization on an array (Jaccoud et al. 2001), it has been adapted to

produce higher density of markers (tens of thousands) by sequencing the representations on a next-generation sequencing machine (Sansaloni et al. 2011). When compared with GBS, DArT-seq produces less markers, but deeper coverage per marker, allowing reliable calling of heterozygote samples.

In the case of maize, large DArTseq datasets from 577 elite CIMMYT maize inbred lines (CMLs) have been generated using a combination of two enzymes, *PstI* and *HPaII* and 96-plex sequencing (Chen et al. 2016). Data is publicly available through CIMMYT Dataverse (https://data.cimmyt.org/dataverse/root?q=CML+dartseq). The generated SNPs were used to evaluate quality control (QC) of the germplasm, to identify mislabeled packages or plots, and to measure heterogeneity of each germplasm entry. The study analyzed several factors affecting the cost of data generation for QC purposes and found that a subset of only 80 markers in a sample of 192 individuals worked well for broad QC of germplasm in the bank (Chen et al. 2016). In addition, the technology has also been used to analyze 447 inbred lines preserved at the germplasm bank from the Federal University of Lavras (Brazil), and the information was used to select resistant varieties to ear rot (dos Santos et al. 2016).

3.3.3 rAmpSeq

Enzyme-based reduction representation libraries require high-quality DNA, which is expensive, laborious, and under certain circumstances complicated to obtain. This is a serious obstacle for genomic selection because thousands of samples need to be genotyped in a very short period at a very low cost per sample. rAmpSeq (Buckler et al. 2016) is an amplicon sequencing technology that allows the use of low-quality DNA, focuses on the repetitive regions of the genome, and produces hundreds to thousands of markers across a genome. Genotyping these regions of the genome is an old approach, but the recent improvement of sequencing technologies, with reads more than 150 bp, has now allowed it to be done at scale. Working with repetitive regions solves several problems that similar technologies like AmpSeq (Yang et al. 2016) have. It reduces PCR competition that causes problems to score heterozygote samples because all amplicons are similar in size and composition, and it only needs a few primer pairs that can be designed without previous knowledge of the sequences to amplify, which reduces bias. On the other hand, when compared with GBS, rAmpSeq produces fewer markers, needs knowledge of a reference genome to initially design the primers, and is focused on intergenic regions, which makes the bioinformatics processing complicated and customized for each species. In addition, the current analysis pipelines can only score dominant markers. The major advantage of the technology is simplicity, which provides a very competitive cost per sample (approximately $5/sample). Cornell University and CIMMYT are currently working together to use this technology for genomic selection in breeding programs serving low- and middle-income countries. Genomic prediction accuracies and cost reports will be available in the near future (Zhang et al. 2017).

3.3.4 Transcriptome Sequencing

Another approach to obtaining a reduced representation of the genome consists of using transcriptome data. The process involves isolating mRNA and converting it into a library of cDNA that is then used as input in the next-generation sequencing machine (Nagalakshmi et al. 2008; Wang et al. 2009). Only about 1–2% of the genome is usually represented with this approach, but these sequences are likely to contain a high number of functional variants (Scheben et al. 2017). This technique provides information about diversity but also expression differences in the same tissue. Its major limitation is that the gene must be expressed in the sampled tissue to obtain data (Hirsch et al. 2014a). RNA-Seq has been used in maize to obtain genetic markers, in combination with other technologies or alone. For example, a modification of a bulk segregant

analysis that made use of RNA-Seq data was used to clone the glossy3 gene in maize (Liu et al. 2012).

Analysis of RNA-Seq data from a diverse set of inbred lines has provided useful information to differentiate between the core and the dispensable portions of the genome (pan-genome). An initial dataset generated with seedlings of 21 diverse inbred lines showed that the transcribed gene set was highly variable with only about 49% of the genes expressed in all the lines. Moreover, this study found 350,710 SNPs, with 22,831 genes having at least one SNP (Hansey et al. 2012). In 2014, a study with a bigger panel of 503 inbred lines at the seedling stage, representative of the major US grain heterotic groups and including all 465 inbred lines in the Wisconsin Diversity Set (Hansey et al. 2011), found that only 16.4% of the genes were expressed in all of them. The study identified more than 1.5 million SNPs in the transcribed genome and a subset of markers with higher quality was used to map flowering time and juvenile-to-adult vegetative phase transition (Hirsch et al. 2014b). This study helped to understand the dynamics of the maize pan-genome and demonstrated that a substantial proportion of the variation in quantitative traits may be due to genes not present in the reference genome. This set of markers for the Wisconsin Diversity Panel is publicly available on Data Dryad (https://doi.org/10.5061/dryad.r73c5).

A variation of the RNA-Seq method, known as 3′ RNA-Seq, consists of sequencing only one fragment per transcript in the 3′ region (Torres et al. 2008). The main advantage of this method when compared with traditional RNA-Seq is that since only one fragment is sequenced per transcript, the level of expression can be estimated directly from the number of reads regardless of transcript length (Tandonnet and Torres 2017). On the other hand, since only a small portion of the transcribed gene is sequenced, the number of markers obtained is very limited. However, the protocol allows cheaper and high-throughput characterization of the transcriptome. 3′ RNA-Seq has recently been used to characterize seven different tissues in approximately 300 maize inbred lines (Flint-Garcia et al. 2005) and demonstrates how rare alleles affect expression and plant phenotypes (Kremling et al. 2018). Public data from this study is available on Panzea (http://www.panzea.org).

3.4 Future Directions and Challenges: Cheap Sequencing and Data Sharing

Over the last decade, an incredible drop in prices, reduction in sequencing errors, and increased read lengths have transformed next-generation sequencing and altered what information can be easily obtained to study the maize genome, its diversity, and its changes trough domestication and breeding. With this revolution of prices, target enrichment or reduction of genome complexity is not a key element for the development of new molecular markers anymore, and whole genome sequencing can become more widespread for crop genotyping (Scheben et al. 2017). Availability of resequencing data from multiple accessions of the same species has allowed a further exploration of the intraspecific variation, and it has become clear that a system to represent that diversity that goes beyond a single reference genome is needed (Hurgobin and Edwards 2017).

In maize, the first-generation haplotype map to study diversity at the whole genome level for 27 diverse inbred lines was published less than 10 years ago, in 2009 (Gore et al. 2009), right after the publication of the first version of the B73 reference genome (Schnable et al. 2009). The study still targeted only the low-repetitive regions of the genome, using a combination of two restriction enzymes that only sampled about 20% of the genome. A study with six elite commercial inbred lines covering the entire genome was published shortly after (Lai et al. 2010). The next version of a haplotype map for maize already included 103 inbred lines across pre-domesticated and domesticated *Zea mays* varieties and covered the entire genome (Chia et al. 2012). Similarly, 278 temperate inbred lines from different stages of breeding history were sequenced (Jiao et al. 2012). The latest version of

the Hapmap of maize combined over 12 trillion base pairs of sequence data, covering 1,218 different taxa (Bukowski et al. 2017). During the same period of time, three additional improved versions of the reference genome (Jiao et al. 2017b) have been published, and technology has enabled the publication of assemblies for another six inbred lines of variable completeness (Lu et al. 2015; Hirsch et al. 2016; Unterseer et al. 2017; Yang et al. 2017) with more efforts currently ongoing (Andorf et al. 2016a). Resequencing technologies have also been recently used to construct high-density recombination maps of important mapping populations, like IBM using deep sequencing of the Mo17 parent and skim of progeny (Liu et al. 2015), or to study the origins, demography, and selection of maize (Hufford et al. 2012b; Brandenburg et al. 2017).

The current technology offers enormous possibilities for obtaining genomic information at many different throughput and price point levels. A simple workflow that allows efficient use of those possibilities to mine and manage diversity and prediction of phenotypes from genotypes while maximizing resource usage is a key. First, resources should be invested in deep exploration, that is, good-quality whole genome sequence data of the most important germplasm. This selection would cover a bigger pool of alleles that researchers and breeders can select for their work and provide a collection of all the genes at the clade level (Varshney et al. 2016). Then, cheaper methods can provide skim or reduced representation sequencing that can be used for maintenance of the banks or construction of high-density linkage maps trough imputation of the haplotype blocks from the whole genome samples. These skim sequencing methods could target regions of interest to guarantee an even distribution of markers and allow breeders to follow their most important genes clearly through their programs. Additionally, these approaches should work on low-quality DNA to reduce costs and allow their use in environments where resources are scarce like developing countries and have quick turnaround time. All these are similar to what single-plex systems currently offer (Chen et al. 2016), but should be implemented in a high-throughput manner (at least a few thousand markers to track recombination breakpoints).

The volume of data being generated by high-throughput technologies will require new tools, methods, and policies for data storage, data processing, and data sharing. Bioinformatics, mostly related to data management and the analysis of a complex and repetitive genome, is still a bottleneck. Specific analysis pipelines for each of the different library construction methods need to be developed each time technology evolves and most of the analysis developed for human genomes do not work well in plant species with more complicated genomes. Generated datasets are then difficult to combine since they do not share most of the markers and analysis biases. With the speed at which technology improvements keep happening, a universal pipeline that can be applied to any plant species, capturing information about the pan-genome and allowing the combination of data from any sequencing method is essential to take maximum advantage of existing and future datasets. In addition, rational agreements to establish policies to allow public data sharing while encouraging a fair use of maize genetic resources and support for data-related initiatives that facilitate the use of the genomic information by breeders like the genomic open-source breeding informatics initiative (GOBii; http://gobiiproject.org/) or community resources like MaizeGDB (Andorf et al. 2016b) are crucial to keep making progress in understanding the links between genotype and phenotype and breeding for a more productive and sustainable maize (Rasheed et al. 2017; Halewood et al. 2018).

In summary, although there are some challenges that need to be addressed during the next few years, a new interesting era to further explore genomic diversity for breeding and genomics through high-throughput and affordable genotyping is ahead of us.

Acknowledgements M. Cinta Romay is supported by NSF Plant Genome Project (NSF #1238014) and USDA-ARS.

References

Andorf CM, Cannon EK, Portwood JL et al (2016a) MaizeGDB update: new tools, data and interface for the maize model organism database. Nucleic Acids Res 44:D1195–D1201. https://doi.org/10.1093/nar/gkv1007

Andorf CM, Cannon EK, Portwood JL et al (2016b) MaizeGDB update: new tools, data and interface for the maize model organism database. Nucleic Acids Res 44:D1195–D1201. https://doi.org/10.1093/nar/gkv1007

Baird NA, Etter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376. https://doi.org/10.1371/journal.pone.0003376

Beissinger TM, Hirsch CN, Sekhon RS et al (2013) Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics 193:1073–1081. https://doi.org/10.1534/genetics.112.147710

Brandenburg J-T, Mary-Huard T, Rigaill G et al (2017) Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLoS Genet 13:e1006666. https://doi.org/10.1371/journal.pgen.1006666

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98:116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Curr Opin Plant Biol 9:172–176. https://doi.org/10.1016/j.pbi.2006.01.013

Buckler ES, Ilut DC, Wang X, et al (2016) rAmpSeq: Using repetitive sequences for robust genotyping. https://doi.org/10.1101/096628

Bukowski R, Guo X, Lu Y et al (2017) Construction of the third generation zea mays haplotype map. Gigascience. https://doi.org/10.1093/gigascience/gix134

Bybee SM, Bracken-Grissom H, Haynes BD et al (2011) Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol Evol 3:1312–1323. https://doi.org/10.1093/gbe/evr106

Chen J, Zavala C, Ortega N et al (2016) The development of quality control genotyping approaches: a case study using elite maize lines. PLoS One 11:e0157236. https://doi.org/10.1371/journal.pone.0157236

Chia J-M, Song C, Bradbury PJ et al (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807. https://doi.org/10.1038/ng.2313

Cook JP, McMullen MD, Holland JB et al (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158:824–834. https://doi.org/10.1104/pp.111.185033

Cui D, Wu D, Somarathna Y et al (2015) QTL mapping for salt tolerance based on snp markers at the seedling stage in maize (Zea mays L.). Euphytica 203:273–283. https://doi.org/10.1007/s10681-014-1250-x

Dell'Acqua M, Gatti DM, Pea G et al (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol 16:167. https://doi.org/10.1186/s13059-015-0716-z

Diversity Arrays Technology Pty. Ltd (2018) DArTseq. http://www.diversityarrays.com/dart-application-dartseq. Accessed 26 Mar 2018

dos Santos JPR, Pires LPM, de Castro Vasconcellos RC et al (2016) Genomic selection to resistance to Stenocarpella maydis in maize lines using DArTseq markers. BMC Genet 17:86. https://doi.org/10.1186/s12863-016-0392-3

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Egan AN, Schlueter J, Spooner DM (2012) Applications of next-generation sequencing in plant biology. Am J Bot 99:175–185. https://doi.org/10.3732/ajb.1200020

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

FAOSTAT (2018) FAOSTAT Data. In: Food Agric. Organ. United Nations, Rome. http://faostat.fao.org. Accessed 22 Mar 2018

Flint-Garcia S, Thuillet A-C, Yu J et al (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578. https://doi.org/10.1073/pnas.132259199

Fu J, Cheng Y, Linghu J et al (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. Nat Commun 4:https://doi.org/10.1038/ncomms3832

Fu Y, Springer NM, Gerhardt DJ et al (2010) Repeat subtraction-mediated sequence capture from a complex genome. Plant J 62:898–909. https://doi.org/10.1111/j.1365-313X.2010.04196.x

Ganal MW, Durstewitz G, Polley A et al (2011) A large maize (Zea mays L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6:e28334. https://doi.org/10.1371/journal.pone.0028334

Ganal MW, Polley A, Graner E-M et al (2012) Large SNP arrays for genotyping in crop plants. J Biosci 37:821–828. https://doi.org/10.1007/s12038-012-9225-3

Giraud H, Bauland C, Falque M et al (2017) Reciprocal genetics: identifying QTL for general and specific combining abilities in hybrids between multiparental

populations from two maize (Zea mays L.) heterotic groups. Genetics 207:1167–1180. https://doi.org/10.1534/genetics.117.300305

Glaubitz JC, Casstevens TM, Lu F et al (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. https://doi.org/10.1371/journal.pone.0090346

Gore MA, Chia J-M, Elshire RJ, et al (2009) A first-generation haplotype map of maize. Science 326 (80):1115–1117. https://doi.org/10.1126/science.1177837

Gouesnard B, Negro S, Laffray A et al (2017) Genotyping-by-sequencing highlights original diversity patterns within a European collection of 1191 maize flint lines, as compared to the maize USDA genebank. Theor Appl Genet 130:2165–2189. https://doi.org/10.1007/s00122-017-2949-6

Guo S, Ku L, Qi J et al (2015) Genetic analysis and major quantitative trait locus mapping of leaf widths at different positions in multiple populations. PLoS One 10:e0119095. https://doi.org/10.1371/journal.pone.0119095

Gupta PK (2008) Single-molecule DNA sequencing technologies for future genomics research. Trends Biotechnol 26:602–611. https://doi.org/10.1016/j.tibtech.2008.07.003

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity (Edinb) 101:5–18. https://doi.org/10.1038/hdy.2008.35

Halewood M, Chiurugwi T, Sackville Hamilton R et al (2018) Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. New Phytol 217:1407–1419. https://doi.org/10.1111/nph.14993

Han S, Miedaner T, Utz HF et al (2018) Genomic prediction and GWAS of gibberella ear rot resistance traits in dent and flint lines of a public maize breeding program. Euphytica 214:6. https://doi.org/10.1007/s10681-017-2090-2

Hansey CN, Johnson JM, Sekhon RS et al (2011) Genetic diversity of a maize association population with restricted phenology. Crop Sci 51:704–715. https://doi.org/10.2135/cropsci2010.03.0178

Hansey CN, Vaillancourt B, Sekhon RS et al (2012) Maize (Zea mays L.) genome diversity as Revealed by RNA-sequencing. PLoS One 7:e33071. https://doi.org/10.1371/journal.pone.0033071

Hao D, Cheng Y, Chen G et al (2015a) Identification of significant single nucleotide polymorphisms for resistance to maize rough dwarf disease in elite maize (Zea mays L.) inbred lines. Euphytica 203:109–120. https://doi.org/10.1007/s10681-014-1277-z

Hao D, Xue L, Yuan J et al (2017) Genetic dissection of starch paste viscosity characteristics in waxy maize revealed by high-density SNPs in a recombinant inbred line population. Mol Breed 37:50. https://doi.org/10.1007/s11032-017-0652-6

Hao D, Zhang Z, Cheng Y et al (2015b) Identification of genetic differentiation between waxy and common maize by SNP genotyping. PLoS One 10:e0142585. https://doi.org/10.1371/journal.pone.0142585

He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay, pp 75–86

Hirsch CD, Evans J, Buell CR, Hirsch CN (2014a) Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. Brief Funct Genomics 13:257–267. https://doi.org/10.1093/bfgp/elt051

Hirsch CN, Foerster JM, Johnson JM et al (2014b) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26:121–135. https://doi.org/10.1105/tpc.113.119982

Hirsch CN, Hirsch CD, Brohammer AB et al (2016) Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. Plant Cell 28:2700–2714. https://doi.org/10.1105/tpc.16.00353

Hodges E, Xuan Z, Balija V et al (2007) Genome-wide in situ exon capture for selective resequencing. Nat Genet 39:1522–1527. https://doi.org/10.1038/ng.2007.42

Hoffmann TJ, Zhan Y, Kvale MN, Hesselson SE, Gollub J, Iribarren C, Lu Y, Mei G, Purdy MM, Quesenberry C, Rowell S, Shapero MH, Smethurst D, Somkin CP, Van den Eeden SK, Walter L, Webster T, Whitmer RA, Finn A, Schaefer C, Kwok P- Y, Risch, N. (2011). Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. Genomics 98 (6): 422–430.

Hu S, Wang C, Sanchez DL, et al (2017) Gibberellins promote brassinosteroids action and both increase heterosis for plant height in maize (Zea mays L.). Front Plant Sci 8.https://doi.org/10.3389/fpls.2017.01039

Hufford MB, Xu X, van Heerwaarden J et al (2012a) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811. https://doi.org/10.1038/ng.2309

Hufford MB, Xu X, van Heerwaarden J et al (2012b) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811. https://doi.org/10.1038/ng.2309

Hurgobin B, Edwards D (2017) SNP discovery using a pangenome: has the single reference approach become obsolete? Biology (Basel) 6:21. https://doi.org/10.3390/biology6010021

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:E25

Jiao Y, Peluso P, Shi J et al (2017a) Improved maize reference genome with single-molecule technologies. Nature. https://doi.org/10.1038/nature22971

Jiao Y, Peluso P, Shi J et al (2017b) Improved maize reference genome with. Nature 546:524–527. https://doi.org/10.1038/nature22971

Jiao Y, Zhao H, Ren L et al (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44:812–815. https://doi.org/10.1038/ng.2312

Jones E, Chu W-C, Ayele M et al (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (Zea mays L.) germplasm. Mol Breed 24:165–176. https://doi.org/10.1007/s11032-009-9281-z

Kahler AL, Kahler JL, Thompson SA et al (2010) North american study on essential derivation in maize: II. Selection and evaluation of a panel of simple sequence repeat loci. Crop Sci 50:486. https://doi.org/10.2135/cropsci2009.03.0121

Kremling KAG, Chen S-Y, Su M-H et al (2018) Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature. https://doi.org/10.1038/nature25966

Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics 2012:1–15. https://doi.org/10.1155/2012/831460

Lai J, Li R, Xu X et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:1027–1030. https://doi.org/10.1038/ng.684

Lee M, Sharopova N, Beavis WD et al (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol Biol 48:453–461

Li H, Peng Z, Yang X et al (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45:43–50. https://doi.org/10.1038/ng.2484

Li K, Yan J, Li J, Yang X (2014) Genetic architecture of rind penetrometer resistance in two maize recombinant inbred line populations. BMC Plant Biol 14:152. https://doi.org/10.1186/1471-2229-14-152

Liu H, Niu Y, Gonzalez-Portilla PJ et al (2015) An ultra-high-density map as a community resource for discerning the genetic basis of quantitative traits in maize. BMC Genom 16:1078. https://doi.org/10.1186/s12864-015-2242-5

Liu S, Yeh C-T, Tang HM et al (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS One 7:e36406. https://doi.org/10.1371/journal.pone.0036406

Lu F, Romay MC, Glaubitz JC, et al (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. Nat Commun 6: https://doi.org/10.1038/ncomms7914

Luo M, Zhao Y, Zhang R et al (2017) Mapping of a major QTL for salt tolerance of mature field-grown maize plants based on SNP markers. BMC Plant Biol 17:140. https://doi.org/10.1186/s12870-017-1090-7

Manching H, Sengupta S, Hopper KR, et al (2017) Phased Genotyping-by-sequencing enhances analysis of genetic diversity and reveals divergent copy number variants in maize. G3 Genes, Genomes, Genet 7:2161–2170. https://doi.org/10.1534/g3.117.042036

Mayer M, Unterseer S, Bauer E et al (2017) Is there an optimum level of diversity in utilization of genetic resources? Theor Appl Genet 130:2283–2295. https://doi.org/10.1007/s00122-017-2959-4

McMullen MD, Kresovich S, Villeda HS, et al (2009) Genetic properties of the maize nested association mapping population. Science 325(80):737–740. https://doi.org/10.1126/science.1174320

Meng Y, Li J, Liu J et al (2016) Ploidy effect and genetic architecture exploration of stalk traits using DH and its corresponding haploid populations in maize. BMC Plant Biol 16:50. https://doi.org/10.1186/s12870-016-0742-3

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Millet E, Welcker C, Kruijer W, et al (2016) Genome-wide analysis of yield in Europe: allelic effects as functions of drought and heat scenarios. Plant Phys 00621.2016. https://doi.org/10.1104/pp.16.00621

Morgante M, Brunner S, Pea G et al (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002. https://doi.org/10.1038/ng1615

Nagalakshmi U, Wang Z, Waern K, et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320(80):1344–1349. https://doi.org/10.1126/science.1158441

Nelson PT, Coles ND, Holland JB et al (2008) Molecular characterization of maize inbreds with expired US plant variety protection. Crop Sci 48:1673–1686. https://doi.org/10.2135/cropsci2008.02.0092

Nelson PT, Krakowsky MD, Coles ND et al (2016) Genetic characterization of the North Carolina State University maize lines. Crop Sci 56:259. https://doi.org/10.2135/cropsci2015.09.0532

Ott A, Liu S, Schnable JC et al (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Res 45:e178–e178. https://doi.org/10.1093/nar/gkx853

Pan Q, Ali F, Yang X et al (2012) Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. PLoS One 7:e52777. https://doi.org/10.1371/journal.pone.0052777

Pestsova E, Lichtblau D, Wever C et al (2016) QTL mapping of seedling root traits associated with nitrogen and water use efficiency in maize. Euphytica 209:585–602. https://doi.org/10.1007/s10681-015-1625-7

Peterson BK, Weber JN, Kay EH et al (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. Genome Biol Evol 5:1594–1609. https://doi.org/10.1093/gbe/evt109

Rabiner LR (1989) A tutorial on hidden markov models and selected applications in speech recognition. Proc IEEE 77:257–286. https://doi.org/10.1109/5.18626

Rasheed A, Hao Y, Xia X et al (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Mol Plant 10:1047–1064. https://doi.org/10.1016/j.molp.2017.06.008

Rincent R, Nicolas S, Bouchet S et al (2014) Dent and flint maize diversity panels reveal important genetic potential for increasing biomass production. Theor Appl Genet 127:2313–2331. https://doi.org/10.1007/s00122-014-2379-7

Romay MC, Millard MJ, Glaubitz JC et al (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14:R55. https://doi.org/10.1186/gb-2013-14-6-r55

Romero Navarro JA, Willcox M, Burgueño J et al (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet 49:476–480. https://doi.org/10.1038/ng.3784

Rousselle Y, Jones E, Charcosset A et al (2015) Study on essential derivation in maize: III. Selection and evaluation of a panel of single nucleotide polymorphism loci for use in European and North American Germplasm. Crop Sci 55:1170. https://doi.org/10.2135/cropsci2014.09.0627

Sansaloni C, Petroli C, Jaccoud D et al (2011) Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of eucalyptus. BMC Proc 5:54. https://doi.org/10.1186/1753-6561-5-S7-P54

Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J 15:149–161. https://doi.org/10.1111/pbi.12645

Schnable PS, Ware D, Fulton RS, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(80):1112–1115. https://doi.org/10.1126/science.1178534

Semagn K, Magorokosho C, Vivek BS et al (2012) Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. BMC Genom 13:113. https://doi.org/10.1186/1471-2164-13-113

Steemers FJ, Gunderson KL (2007) Whole genome genotyping technologies on the BeadArrayTM platform.Biotechnol J 2:41–49. https://doi.org/10.1002/biot.200600213

Shen R, Fan J-B, Campbell D et al (2005) High-throughput SNP genotyping on universal bead arrays. Mutat Res Mol Mech Mutagen 573:70–82. https://doi.org/10.1016/j.mrfmmm.2004.07.022

Shi Y, LI G, TIAN Z, et al (2016) Genetic dissection of seed vigour traits in maize (Zea mays L.) under low-temperature conditions. J Genet 95:1017–1022. https://doi.org/10.1007/s12041-016-0714-2

Song W, Shi Z, Xing J et al (2017) Molecular mapping of quantitative trait loci for grain moisture at harvest in maize. Plant Breed 136:28–32. https://doi.org/10.1111/pbr.12430

Springer NM, Ying K, Fu Y et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5:e1000734. https://doi.org/10.1371/journal.pgen.1000734

Stuber C. W, Goodman MM (1983) Allozyme genotypes for popular and historically important inbred lines of corn, Zea mays L. USDA Agric Res Results, South Ser 16

Swarts K, Gutaker RM, Benz B, et al (2017) Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science 357 (80). https://doi.org/10.1126/science.aam9425

Swarts K, Li H, Romero Navarro JA, et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7. https://doi.org/10.3835/plantgenome2014.05.0023

Tandonnet S, Torres TT (2017) Traditional versus 3′ RNA-seq in a non-model species. Genomics Data 11:9–16. https://doi.org/10.1016/j.gdata.2016.11.002

Tian H-L, Wang F-G, Zhao J-R et al (2015) Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of chinese maize varieties. Mol Breed 35:136. https://doi.org/10.1007/s11032-015-0335-0

Tobler AR, Short S, Andersen MR et al (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. J Biomol Tech 16:398–406

Torres TT, Metta M, Ottenwälder B, Schlötterer C (2008) Gene expression profiling by massively parallel sequencing. Genome Res 18:172–177. https://doi.org/10.1101/gr.6984908

Unterseer S, Bauer E, Haberer G et al (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genom 15:823. https://doi.org/10.1186/1471-2164-15-823

Unterseer S, Seidel MA, Bauer E, et al (2017) European flint reference sequences complement the maize pan-genome. bioRxiv

van Heerwaarden J, Hufford MB, Ross-Ibarra J (2012) Historical genomics of North American maize. Proc Natl Acad Sci 109:12420–12425. https://doi.org/10.1073/pnas.1209275109

Varshney RK, Singh VK, Hickey JM et al (2016) Analytical and decision support tools for genomics-assisted breeding. Trends Plant Sci 21:354–363. https://doi.org/10.1016/j.tplants.2015.10.018

Wallace JG, Bradbury PJ, Zhang N et al (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet 10:e1004845. https://doi.org/10.1371/journal.pgen.1004845

Wang C, Hu S, Gardner C, Lübberstedt T (2017) Emerging avenues for utilization of exotic germplasm. Trends Plant Sci 22:624–637. https://doi.org/10.1016/j.tplants.2017.04.002

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Methods 9:808–810. https://doi.org/10.1038/nmeth.2023

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Wen W, Araus JL, Shah T et al (2011a) Molecular characterization of a diverse maize inbred line collection and its potential utilization for stress tolerance improvement. Crop Sci 51:2569–2581. https://doi.org/10.2135/cropsci2010.08.0465

Wen W, Taba S, Shah T et al (2011b) Detection of genetic integrity of conserved maize (Zea mays L.) germplasm in genebanks using SNP markers. Genet Resour Crop Evol 58:189–207. https://doi.org/10.1007/s10722-010-9562-8

Wu Y, San Vicente F, Huang K et al (2016) Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theor Appl Genet 129:753–765. https://doi.org/10.1007/s00122-016-2664-8

Xu C, Ren Y, Jian Y et al (2017a) Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. Mol Breed 37:20. https://doi.org/10.1007/s11032-017-0622-z

Xu Y, Xu C, Xu S (2017b) Prediction and association mapping of agronomic traits in maize using multiple omic data. Heredity (Edinb) 1–11. https://doi.org/10.1038/hdy.2017.27

Yan J, Shah T, Warburton ML et al (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4:e8451. https://doi.org/10.1371/journal.pone.0008451

Yan J, Yang X, Shah T et al (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451. https://doi.org/10.1007/s11032-009-9343-2

Yang J, Jiang H, Yeh C-T et al (2015) Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. Plant J 84:587–596. https://doi.org/10.1111/tpj.13029

Yang N, Xu X-W, Wang R-R et al (2017) Contributions of zea mays subspecies mexicana haplotypes to modern maize. Nat Commun 8:1874. https://doi.org/10.1038/s41467-017-02063-5

Yang S, Fresnedo-Ramírez J, Wang M et al (2016) A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker-assisted selection in grapevine. Hortic Res 3:16002. https://doi.org/10.1038/hortres.2016.2

Yang X, Gao S, Xu S et al (2010) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Mol Breed 28:511–526. https://doi.org/10.1007/s11032-010-9500-7

Yin Z, Wang Y, Wu F et al (2014) Quantitative trait locus mapping of resistance to aspergillus flavus infection using a recombinant inbred line population in maize. Mol Breed 33:39–49. https://doi.org/10.1007/s11032-013-9932-y

Zhang A, Wang H, Beyene Y et al (2017) Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. Front Plant Sci 8. https://doi.org/10.3389/fpls.2017.01916

Zhou G, Hao D, Chen G et al (2016) Genome-wide association study of the husk number and weight in maize (Zea mays L.). Euphytica 210:195–205. https://doi.org/10.1007/s10681-016-1698-y

Zhou Q, Dong Y, Shi Q et al (2017) Verification and fine mapping of qGW1.05, a major QTL for grain weight in maize (Zea mays L.). Mol Genet Genomics 292:871–881. https://doi.org/10.1007/s00438-017-1318-0

Part II
Genome Structure and Phenomena

Maize Transposable Element Dynamics

4

Jeffrey L. Bennetzen

Abstract

Transposable elements (TEs) are among the most important factors in the evolution of gene and genome structure/function in plants. All plant genomes contain mostly quiescent TEs that are activated, independently by family, in currently unpredictable timeframes by largely unknown phenomena. Different reawakened or horizontally transferred TE families can remain active for as little as a few years to as much as a few million years, and the reasons for these duration-of-activity differences are also not known. The maize lineage has seen extraordinary TE activity, and changes in TE activity, over the last few million years, and much of this dynamic continues to be ongoing. Hence, studies of TE biology have been particularly informative in maize, and will continue to be so. This review describes the history of TE activity over the last few million years in the maize lineage, briefly mentions the extensive literature regarding maize TE regulation, and suggests approaches for characterizing the processes that determine which TEs are active: where, when, how and why.

J. L. Bennetzen (✉)
Department of Genetics, University of Georgia, 30602 Athens, GA, USA
e-mail: maize@uga.edu

4.1 Introduction

Transposable elements (TEs) were first discovered and described in maize, initially through the brilliant studies of genome dynamics conducted by McClintock (McClintock 1951, 1953). Although first considered by many to be a technical artifact or a maize-specific oddity, subsequent research has shown that TE activity has been present in all biological lineages and continues to play leading roles in the processes of gene and genome evolution. Despite TE ubiquity across genomes, and genomic evidence of recent (within the last few million years) activity in the great majority of taxa of both plants and animals, it has been surprisingly difficult to find currently active TEs. In flowering plants (angiosperms), only a handful of species have been associated with TEs with ongoing transposition that can be studied in real-time. Often, these activities were only identified after a severe stress was employed to reactivate epigenetically silenced TEs (Peschke et al. 1987; Grandbastien et al. 1989; Peterson 1991; Hirochika et al. 1996). The maize genome, for reasons unknown, provides an exception to this rarity of current TE activity. At least ten TE families have been associated with mutations that occurred in the last few years in maize. Despite this, most maize TE families continue to be inactive in most maize lineages, such that, even for this "most TE-unstable genome" in the eukaryotic world, silenced TEs are

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_4

still the general rule. This chapter discusses the history of TE activity in the maize lineage (*Zea*) and also presents arguments regarding the transitions between TE active and TE inactive states.

4.2 The Recent History of TE Activity in the *Zea* Lineage

4.2.1 Panicoid Grass Genomes

Early molecular analyses of the maize nuclear genome indicated that it was >70% TEs (SanMiguel et al. 1996), a value that was updated to >85% upon full-genome sequence analysis (Schnable et al. 2009). Inspection of the maize genome sequence led to the conclusion that there are ~210,000 TE-encoded genes compared to the ~35,000 protein-encoding nuclear genes that are not TE-derived (Bennetzen 2009). Moreover, transcripts homologous to TEs make up a minimal estimate of 0.1–7.7% of the polyadenylated mRNA in investigated maize tissues (Vicient 2010), so one might expect that TE activity would routinely be observed, even on a daily basis. As in other plants, however, most maize lines do not have active TEs that can be detected by transpositional, mutational or somatic instability assays. Hence, the very abundant TEs in all plant species are mostly quiescent at any given time, with activity that can be measured in an evolutionary timeframe rather than in real-time experiments.

The first analysis to try to determine the evolutionary timing of TE activity in plants was conducted on the long terminal repeat (LTR) retrotransposons of maize (SanMiguel et al. 1998). This study indicated that >90% of current TEs had to have transposed into their current locations within the last six million years, primarily within the last 1–2 million years. This recency of TE activity has also been observed in almost all other plant genome investigated with the notable exception of gymnosperms (Nystedt et al. 2013) and perhaps other very large (>10 GB) genomes (Kelly et al. 2015). The high frequency of TE transposition, mostly LTR retrotransposons, was shown to be the leading cause of genome size expansion (reviewed in Bennetzen et al. 2005), but it was not initially clear why most of these expansions seemed to have originated from recent TE amplifications rather than an accumulated history of ancient TE amplifications, as observed in gymnosperms, *Fritillaria*, olive (Barghini et al. 2015) and a few other species. However, the discovery of rapid processes for DNA removal, initially in Arabidopsis and rice (Devos et al. 2002; Ma et al. 2004), indicated that old TEs were missed only because they had been degraded beyond recognition (Maumus and Quesneville 2014). Most of this degradation was caused by the accumulation of small deletions, primarily via illegitimate recombination outcomes that are apparently derived from inaccuracies in double-strand break repair (Puchta 2005). Different species varied not only in their frequency and degree of TE amplification bursts, but also in their rate of DNA removal (Vitte and Bennetzen 2006; Puchta 2005; Nystedt et al. 2013; Cossu et al. 2017), thus creating the very dynamic differences observed in plant genome sizes, TE content and genome structure.

In an analysis of the panicoid grass lineage, which includes maize, sorghum, sugarcane and several millets, commonality in the nature of recent TE amplification bursts was found to be shared by some but not all lineages in the ~25 million years of panicoid grass evolution (Estep et al. 2013). For instance, the family *Ji/Opie* was separately active in each of the *Zea* lineages investigated, but not in sorghum or any other grass looked at in this study. In contrast, family *Huck* exhibited recent high activity in all of the investigated panicoids, including maize, sorghum, sugarcane and pearl millet, but not in rice, a non-panicoid grass. This indicated that transpositions were generated for specific TE families in particular time windows, and that other TE families were not detectably activated by these same events (Estep et al. 2013). What phenomenon might have created the potential for convergent activity bursts for specific TE families some >20 million years later is not known.

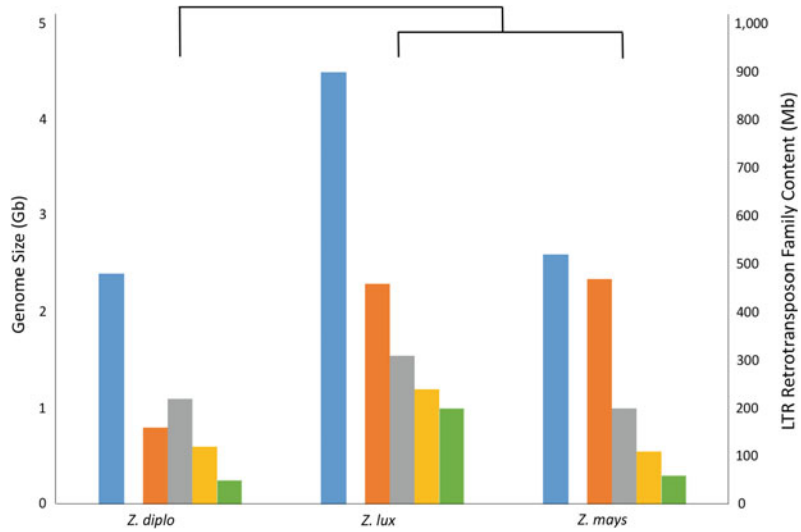

Fig. 4.1 Genome sizes and quantities of the most abundant repeats in three diploid species in the genus *Zea*. Blue indicates genome size (left y axis), while orange, gray, yellow and green indicate the *Huck*, *Ji*, *Opie* and *Prem* LTR retrotransposons, respectively, quantified on the right y axis. The relatedness of the three species is shown by the minimalist phylogenetic tree at the top of the figure. These results are taken from Estep and coworkers (Estep et al. 2013). *Z. dip* stands for *Zea diploperennis* and *Z. lux* stands for *Zea luxurians*. Note that *Z. luxurians* has had numerous LTR retrotransposon amplifications, one of which (*Huck*) was also observed in *Z. mays*

4.2.2 Three Diploid *Zea* Species

The article by Estep and coworkers (Estep et al. 2013) was particularly informative with respect to recent TE activity in members of the genus *Zea*. Two diploid species, *Z. diploperennis* and *Z. mays*, share similar genome sizes of ~2.5 GB, but the diploid *Z. luxurians* has a genome of ~4.5 GB. In less than two million years since these three shared a common ancestor, the *Z. luxurians* genome has nearly doubled in size, mainly caused by the amplification of a few families of LTR retrotransposons. Some of these amplifications were shared by the other two *Zea* species, and some were not (Fig. 4.1). This relative addition of ~2 GB of TEs into *Z. luxurians* is likely to be a huge underestimate of TE activity during that time frame, however, because the processes of DNA removal in the *Zea* lineage are so rapid that maize haplotypes have essentially erased previous intergenic TE content within less than 2 million years (Wang and Dooner 2006). Hence, one can predict that, in the 1–2 million years or less of haplotype divergence between *Z. mays* and *Z. luxurians*, the *Z. luxurians* genome probably acquired ~6 GB of new TEs, but lost ~3 GB of that new DNA to random deletions. *Z. mays*, in the meantime, would have gained ~2 GB of new TEs, and lost ~1 GB of that to deletions, so that the relative growth of *Z. luxurians* would be ~2 GB more than *Z. mays*. The idea that >5 GB of new TE insertions can be generated, in as little as a million years, provides an eye-opening perspective on how dynamic some plant genome lineages can be.

4.2.3 Vertical Versus Horizontal TE Transmission

Most transposable elements are transmitted vertically, through the standard sexual transmission process, as indicated by the fact that careful TE family phylogenies routinely shadow the species phylogenies determined by comparisons of gene sequences. In some animals, however, horizontal transfer of TEs has been observed many times

(reviewed in Schaack et al. 2010). In plants, similar observations of horizontal TE transfer have been few (Diao et al. 2005; El Baidouri et al. 2014), but this may be more an outcome of the small number of studies that have been undertaken rather than any extreme rarity of the phenomenon (Christin et al. 2012; Mahelka et al. 2017).

Transgenic experiments have allowed horizontal transfer to be engineered by interested investigators, and have often led to highly active TEs that can be studied for de novo activity and used for transposon tagging (Baker et al. 1986; Masson and Fedoroff 1989; Lucas et al. 1995). If a TE is moved into a new genome, then its negative regulation by any cytoplasmic factors or nuclear epigenetic silencing might be escaped, thus allowing a burst of activity. Hence, even if rare, horizontal TE transfer should be considered a likely source of active TEs over an evolutionary time frame.

4.3 The Transition Between Inactive and Active TE States

4.3.1 Epigenetic and Genetic Regulation

We have much better understanding of TE silencing than we do of TE activation in plants. Transposable elements were shown to be silenced by DNA methylation-associated processes (Bennetzen 1985; Walbot et al. 1985) even before this had been shown to be true for regular genes. It is now clear that the most abundant small RNAs in large plant genomes like maize, the 24 ntd hc-siRNAs (heterochromatic small interfering RNAs), are derived from TE transcription, processing and DNA methylation by the RdDM (de novo RNA-directed DNA methylation) complex that leads to TE silencing (reviewed in Hammond et al. 2018).

However, it is not at all clear how some TEs can retain activity in a genome where the great majority are silenced. Studies with *Mutator* of maize have shown that the presence of an active transposase can prevent or even reverse epigenetic silencing (Lisch et al. 1995), perhaps by transposase binding to the transposon, thereby prohibiting access of the silencing machinery during somatic or germinal DNA replication. The work of Lisch and collaborators has shown very clearly how a *Mutator* TE with internal inverted duplication can induce a silencing phenomenon across homologous members of the TE family (Slotkin et al. 2005), an idea they propose could be a general phenomenon because TE insertions into members of their own family could also create such inverted structures. This would suggest a dosage-effect outcome, in the sense that the more TEs amplified, then the greater chance that such an inverted structure would be randomly generated.

From a TE perspective, unregulated activity in a genome might not be purely beneficial. Too many mutational insertions, or too great of an occupation of transcriptional and translation machinery, might debilitate the plant in which the TEs are found. Hence, the fascinating endosperm: embryo interactions that are proposed to reinforce epigenetic silencing each plant generation (Martienssen 2010). Some TEs, including in plants (McClintock 1951), exhibit negative self-regulation. Moreover, TEs that preferentially insert into silenced regions of a genome (Gai and Voytas 1998; Baucom et al. 2009b) might also accomplish a moderation of activity simply by increasing their likelihood of silencing by adjacent heterochromatin (Matzke and Matzke 1998; Eichten et al. 2012; Bennetzen and Wang 2018). In this regard, it is interesting that episodes of diversifying selection that may distinguish which LTR retrotransposon families are active and which are not in rice were associated both with reverse transcriptase and integrase. The latter result suggests that finding a new insertion site might be important for initiating or (more likely) sustaining a transposition burst (Baucom et al. 2009a).

4.3.2 Mutational Control

Any sequence in a genome can mutate to an inactive form, and this is likely to happen very

rapidly in a genome like maize that has exceptionally aggressive processes for DNA removal. Genes usually survive this removal process, primarily because of selection for their retention, but perhaps also by more rapid/accurate DNA repair targeted on active regions of the genome (Bohr et al. 1985). Most TE insertions are likely to be selectively neutral, or have negative effects caused by their propensity for mutating genes. Morever, the higher DNA methylation levels in TEs should increase the frequency of C to T transitions (SanMiguel et al. 1998), thus making these mobile DNAs more susceptible to random mutational inactivation. Primarily, because of the absence of selection for their retention, most or all TE families will be lost to the point of complete extinction very rapidly in a genome like maize. Only the random expression of transpositional activity, and the occasional horizontal transfer, have any chance of overcoming this great tide of mutational erosion into extinction.

From the opposite perspective, mutation has the potential to reactivate a quiescent TE, perhaps by creating a sequence variation that is not silenced (e.g., by escaping hc-siRNA homology) or by creating a wholly new TE chimera through an unequal or illegitimate recombination process (Sharma et al. 2008). Horizontal TE transfer, wide crosses leading to polyploidy, TE insertion into TEs, genomic rearrangement (deletions, insertions, inversions, translocations) and TE fragment acquisition by other TEs would all be expected to create opportunities for chimeric TE production that might yield a new mobile DNA able to initially escape epigenetic silencing.

4.3.3 Stress-Induced Activation

From the first observation of *Spm/En* activity arising in maize seed exposed to the Bikini Atoll nuclear bomb tests (Peterson 1991), many cases of TE activation by "genomic stress" have been reported. Taking somatic cells through tissue culture or tissue wounding are routine activational cues for some TE families (Peschke et al. 1987; Grandbastien et al. 1989; Hirochika et al. 1996). The precise mechanism or mechanisms of this activation are not known, although it seems likely that high levels of DNA damage could induce repair events that would play havoc with the maintenance of epigenetic silencing. Some cases of wide crosses, like those that sometimes lead to allopolyploidy, have also been proposed to transiently activate TEs (reviewed in Parisod et al. 2010; Vicient and Casacuberta 2017). Whatever the cause, it is clear that the evolutionary history of each TE family is unique (Wicker and Keller 2007; Baucom et al. 2009b; El Baidouri and Panaud 2013), so the transition from quiescence to active appears to occur on a family-by-family basis, not with an overall genomic activation that becomes fully unconstrained for all TE families at once. This is undoubtedly a good thing, given that tens of thousands of TEs all active in a single nucleus would lead to so many insertional mutations and chromosomal rearrangements that the plant would have zero chance of survival. Moreover, if this severe "genomic shock" (McClintock 1984) had actually saved any plant lineage under severe stress from extinction, then we would perceive this as massively rearranged genomes when comparing very close relatives on a phylogenetic tree. I know of no case where this has been observed, particularly considering that even the massively unstable maize genome maintains excellent genetic collinearity and gene commonality with its close relative sorghum (Bennetzen and Freeling 1993; Tikhonov et al. 1999; Devos and Gale 2000).

4.4 Selection for or Against General TE Activity

The origin of the variation in eukaryotic genome sizes was resolved by the twin discoveries that TEs made up most of the higher eukaryotic genome and that TE content can vary enormously even between closely related species (SanMiguel et al. 1996; Tikhonov et al. 1999; Estep et al. 2013). The question that remains on this issue is: Why does TE content vary so greatly? One simple answer is that TEs persist purely because of their ability to amplify, thereby

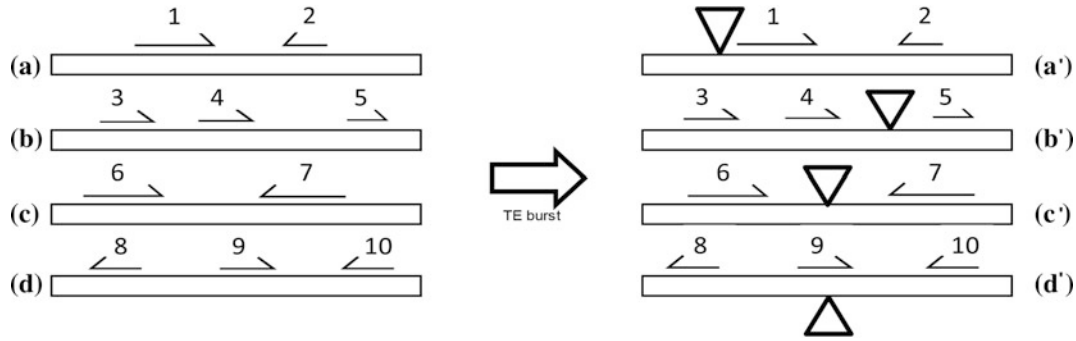

Fig. 4.2 Possible outcomes of a burst of TE amplification. **a**–**d** represent chromosome segments before the TE burst, and **a′**–**d′** represent those same segments after the burst. Genes, their transcript size and their orientation of transcription are indicated by arrows with numbers above them. TE insertions are depicted by triangles, with their points of insertion indicated by where the triangle touches the bar that represents the genomic segment

providing a perfect example of the selfish DNA hypothesis (Doolittle and Sapienza 1980; Orgel and Crick 1980). Selfish DNA is a simple extrapolation of the concept of Darwinian natural selection, but it also demands that this selfish DNA be subject to selection for a possible acquired (gain-of-function) use by the nuclear genome. In fact, there are numerous cases of known advantageous mutations caused by TE insertions. Figure 4.2 depicts a burst of TE insertions that can have positive or negative effects on the genes in the four chromosomal segments shown, a–d. The insertion in a′ could create new promoter regulation in gene 1, and perhaps also in gene 5 in b′. If this is a single active TE family burst, then it is likely that the promoter regulatory sequences provided to genes 1 and 5 by adjacent TEs would have the same gain-of-function outcomes, so that the two unlinked genes now become regulated by the same new environmental or developmental signals. This has been observed in rice (Naito et al. 2009), wherein a suite of genes suddenly became regulated by cold treatment after a burst of *mping* transpositions into promoter regions. Insertions like those near genes 1, 5 and 9 could also inactivate genes by simple structural interruption. Alternatively, the 1, 5 or 9 insertions (especially if the 9 insertion was into an intron) could bring the gene under epigenetic control. In fact, all cases to date of an epigenetic control of a regular plant gene have been derived from an initial TE insertion that created the epigenetically regulated locus (Lisch and Bennetzen 2011). Finally, an insertion like that seen in segment c′ might not have any initial effects on plant fitness, but subsequent mutations (e.g., unequal recombinations and/or deletions causing fusion with a structural gene or its regulatory regions) could lead to evolved value. Because intact TEs and TE fragments make up the majority of most plant genomes (Maumus and Quesneville 2014), they can serve as the raw material for the evolution of all kinds of genomic novelty.

One particular aspect of TE dynamics that deserves a great deal of additional attention in plants is the propensity of TEs for creating new genes. Many studies have shown that all or virtually all plant TE types can incorporate nuclear genes and/or gene fragments within their transpositional modules. This is particularly frequent with Pack-MULEs (Jiang et al. 2004a, b) and *Helitrons* (Morgante et al. 2005), but also common with some LTR retrotransposons. Some genes or acquired fragments inside LTR retrotransposons have exhibited a new function for a TE-amplified gene copy (Kim et al. 2017). For the *Helitrons* of maize, ~4% of the acquired fragments were found to be under purifying selection and ~4% under diversifying selection, suggesting an evolving role for these acquired DNA segments (Yang and Bennetzen 2009). Figure 4.3 shows the possible evolutionary fates of a *Helitron* that sequentially acquired two gene

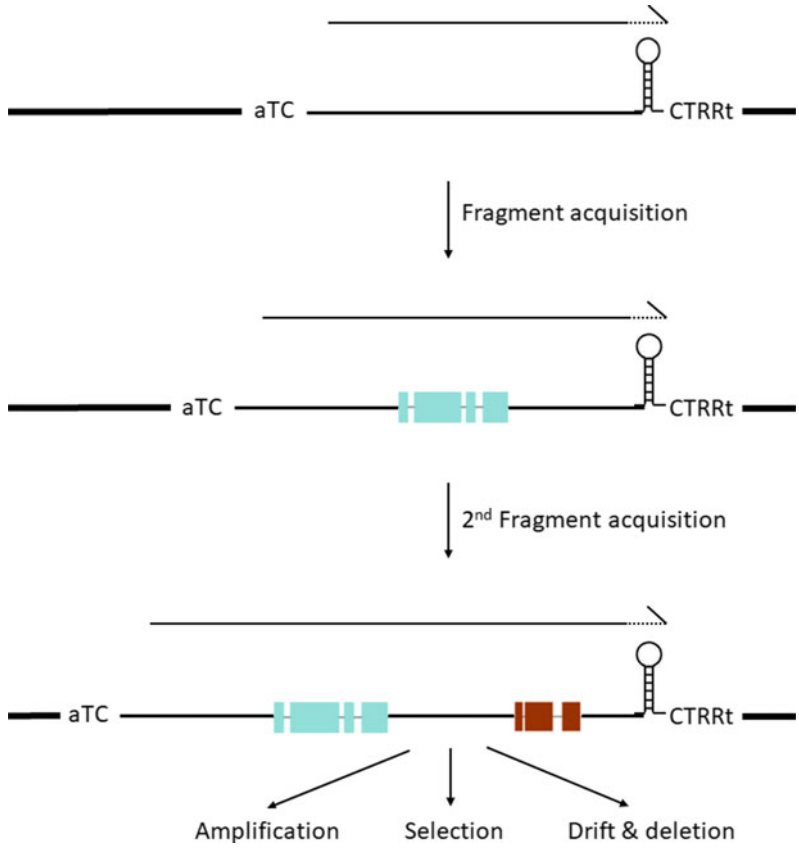

Fig. 4.3 Possible fates of a *Helitron* after it has acquired two gene fragments. *Helitrons* almost always insert within an AT dinucleotide, which is shown by "a" and "t" in this figure. The TC at the 5′ end (relative to transcription from the *Helitron* promoter), and the stem loop followed by CTRR at the 3′ end are standard structural features of *Helitrons*, which do not cause any target site duplication when they insert. This *Helitron* first acquires one gene fragment (blue) and then another (brown), probably during transposition events, so the three *Helitrons* shown here are proposed to have parent: progeny: progeny relationships (top to bottom), but would not be at the same chromosomal locations. Within the acquired gene fragments, the thick regions indicate exons and the thin regions indicate introns. The arrows above the *Helitrons* represent predicted transcripts that could fuse these two acquired gene fragments into a chimeric gene. The arrows at the bottom of the figure indicate possible fates of the *Helitron* with two internal gene fragments. Further amplification may occur (possibly including additional fragment acquisitions), selection against (or very rarely, for) the newly created TE may occur if it affects gene or genome function, or (the most likely outcome) additional mutations will occur. Many of these mutations will be deletions, such that the newly created TE and its chimeric internal acquisitions are lost

fragments, leading to the possible creation of a new gene from fusion of two previously unrelated loci. This process would be an ongoing mechanism of "exon shuffling" (Gilbert 1978) to create genetic novelty. Because maize and other plants have thousands of gene fragments inside TEs in a single nucleus, the opportunity for generating new genes is enormous.

Perhaps this potential for creating new genes and thus new genetic functions is one reason that TEs are so abundant in so many genomes. Given the rate of mutational loss predicted for TEs in most angiosperms, it is difficult to see how any function that is not used at least once every few generations can be maintained. Perhaps the sheer abundance of TEs in a genome means that chance

will allow a few individuals to escape mutational inactivation over any given time frame. If this dynamic holds true, then one expects that some lineages of plants might have more need for TE activity than would others. That is, it could be that plants that are subject to a greater need for genetic innovation will thus have selection for a higher TE activity over evolutionary time (concept reviewed in Bennetzen and Wang 2014). If this is the case, then it seems likely that such TE-rich genomes as maize must belong to lineages that are particularly in need of conservation for an elevated TE activity. This then suggests an "evolutionary genomic shock dependence" model that really only differs from the "genomic shock" model (McClintock 1984) in the timing, duration and breadth of TE activity.

4.5 Conclusions

Transposable element abundance, variation and involvement in gene/genome evolution have been documented for more than 60 years. However, we still do not fully understand why some lineages show a great deal more TE activity than others or why each TE family exhibits a unique history of activity. Maize provides an excellent resource for studying TE dynamics because of its uniquely broad spectrum of currently active TEs and because of a very recent history of extreme TE activity. It would be highly appropriate if future studies undertook examination of processes leading to the activation of quiescent TEs under real-world scenarios of stress exposure or wide crosses. Transgenics, reverse genetics and Mendelian genetics could be used to investigate possible selective advantages of different levels of TE activity in different genetic backgrounds. Maize continues to be a leading organism for the investigation of TE biology and genomic instability, a prominence that is not likely to wain in the foreseeable future.

Acknowledgements The author thanks Aye Htun for her assistance with figures. The writing of this manuscript was supported by the Giles Professorship Endowment at the University of Georgia.

References

Baker B, Schell J, Lörz H, Fedoroff N (1986) Transposition of the maize controlling element "Activator" in tobacco. Proc Natl Acad Sci USA 83:4844–4848

Barghini E, Natali L, Giordani T et al (2015) LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. DNA Res 22:91–100

Baucom RS, Estill JC, Leebens-Mack J, Bennetzen JL (2009a) Natural selection on gene function drives the evolution of LTR retrotransposon families in the rice genome. Genome Res 19:243–254

Baucom RS, Estill JC, Upshaw N et al (2009b) Exceptional diversity, non-random distribution and rapid evolution of retroelements in the B73 maize genome. PLoS Genet 5:e1000732

Bennetzen JL (1985) The regulation of *Mutator* function and *Mu1* transposition. UCLA Symp Mol Cell Biol 35:343–354

Bennetzen JL (2009) Maize genome structure and evolution. In: Bennetzen JL, Hake S (eds) The maize handbook—volume II: genetics and genomics. Springer, New York, pp 179–200

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, collinearity and complementarity. Trends Genet 9:259–261

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95:127–132

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function and evolution of plant genomes. Ann Rev Plant Biol 65:505–530

Bennetzen JL, Wang X (2018) Relationships between gene structure and genome instability in flowering plants. Mol Plant. https://doi.org/10.1016/jmolp.2018.02.003

Bohr VA, Smith CA, Okumoto DS, Hanawalt PC (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. Cell 40:359–369

Christin PA, Edwards EJ, Besnard G et al (2012) Adaptive evolution of C4 photosynthesis through recurrent lateral gene transfer. Curr Biol 22:445–449

Cossu RM, Casola C, Giacomello S et al (2017) LTR retrotransposons show low levels of unequal recombination and high rates of intraelement gene conversion in large plant genomes. Genome Biol Evol 9:3349–3462

Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12:637–646

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601

Diao X, Freeling M, Lisch D (2005) Horizontal transfer of a plant transposon. PLoS Biol 4:e5

Eichten SR, Ellis NA, Makarevitch I et al (2012) Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS Genet 8: e1003127

El Baidouri M, Carpentier MC, Cooke R et al (2014) Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res 24:831–838

El Baidouri M, Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biol Evol 5:954–965

Estep MC, DeBarry JD, Bennetzen JL (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity 110:194–204

Gai XW, Voytas DF (1998) A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell 1:1051–1055

Gilbert W (1978) Why genes in pieces? Nature 271:501

Grandbastien M-A, Spielmann A, Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376–380

Hammond R, Teng C, Meyers BC (2018) Maize small RNAs as seeds of change and stability in gene expression and genome stability. In: The maize genome. Springer, this volume

Hirochika H, Sugimoto K, Otsuki Y et al (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93:7783–7788

Jiang N, Bao Z, Zhang X et al (2004a) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004b) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573

Kelly LJ, Renny-Byfield S, Pellicer J et al (2015) Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol 208:596–607

Kim S, Park J, Yeom SI et al (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol 18:210

Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. Curr Opin Plant Biol 14:156–161

Lisch D, Chomet P, Freeling M (1995) Genetic characterization of the Mutator system in maize: behavior and regulation of Mu transposons in a minimal line. Genetics 139:1777–1796

Lucas H, Feuerbach F, Kunert K et al (1995) RNA-mediated transposition of the tobacco retrotransposon Tnt1 in Arabidopsis thaliana. EMBO J 14:2364–2373

Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869

Mahelka V, Krak K, Kopecký D et al (2017) Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. Proc Natl Acad Sci USA 114:1726–1731

Martienssen RA (2010) Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of Arabidopsis. New Phytol 186:46–53

Masson P, Fedoroff N (1989) Mobility of the maize suppressor-mutator element in transgenic tobacco cells. Proc Natl Acad Sci USA 86:2219–2223

Matzke AJM, Matzke MA (1998) Position effects and epigenetic silencing of plant transgenes. Curr Opin Plant Biol 1:142–148

Maumus F, Quesneville H (2014) Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS One 9:e94101

McClintock B (1951) Chromosome organization and genic expression. Cold Spring Harb Symp 16:13–47

McClintock B (1953) Induction of instability at selected loci in maize. Genetics 38:579–599

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Morgante M, Brunner S, Pea G et al (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Naito K, Zhang F, Tsukiyama T et al (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461:1130–1134

Nystedt B, Street NR, Wetterbom A et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604

Parisod C, Alix K, Just J, Petit M, Sarilar V (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186:37–45

Peschke VM, Phillips R, Gengenbach BG (1987) Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science 238:804–807

Peterson PA (1991) The transposable element-*En*-four decades after Bikini. Genetica 84:63–72

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56:1–14

SanMiguel P, Tikhonov A, Jin Y-K et al (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

SanMiguel P, Gaut BS, Tikhonov A et al (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45

Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537–546

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity and dynamics. Science 326:1112–1115

Sharma A, Schneider KL, Presting GG (2008) Sustained retrotransposition is mediated by nucleotide deletions and interelement recombinations. Proc Natl Acad Sci USA 105:15470–15474

Slotkin RK, Freeling M, Lisch D (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nat Genet 37:641–644

Tikhonov AP, SanMiguel PJ, Nakajima Y et al (1999) Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci USA 96:7409–7414

Vicient CM (2010) Transcriptional activity of transposable elements in maize. BMC Genom 11:601

Vicient CM, Casacuberta JM (2017) Impact of transposable elements on polyploid plant genomes. Ann Bot 120:195–207

Vitte C, Bennetzen JL (2006) Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci USA 103:17638–17643

Walbot V, Chandler C, Taylor L (1985) Alterations in the *Mutator* transposable element family of *Zea mays*. UCLA Symp Mol Cell Biol 35:333–342

Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. Proc Natl Acad Sci USA 103:17644–17649

Wicker T, Keller B (2007) Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. Genome Res 17:1072–1081

Yang L, Bennetzen JL (2009) Distribution, diversity, evolution, and survival of *Helitrons* in the maize genome. Proc Natl Acad Sci USA 106:19922–19927

Genomics of Maize Centromeres

5

Jonathan I. Gent, Natalie J. Nannas, Yalin Liu, Handong Su, Hainan Zhao, Zhi Gao, R. Kelly Dawe, Jiming Jiang, Fangpu Han and James A. Birchler

Abstract

Maize is a model organism for centromere research in part because many of its centromeres are composed of complex sets of genetic elements rather than being dominated by simple tandem repeats common at the centromeres of other taxa. Centromeres in maize range in size to about 2 MB on ~200 MB chromosomes and are characterized by the presence of two repetitive elements: *CentC* is a 156 bp satellite present in highly repetitive arrays, and CRM is an active retrotransposon that apparently prefers centromeric chromatin as sites of insertion. However, there is significant polymorphism for the exact positioning of the centromeric-specific histone, CENH3. Such centromere repositioning events indicate centromeric inactivation and *de novo* formation in maize, both of which have been observed experimentally. Further, *de novo* centromere formation over unique DNA that lacks *CentC* and CRM has been found on chromosomal fragments produced in a variety of ways, sometimes in conjunction with centromere inactivation. The centromere of the supernumerary B chromosome has a specific repetitive sequence interspersed and surrounding the CENH3-enriched core region. This feature has allowed a detailed analysis of the B centromere and the classical phenomenon of centromere misdivision in the background of intact centromeres on A chromosomes. Here we review the DNA and protein components of maize centromeres and how they are maintained for fidelity of chromosome transmission while being malleable in the contexts of both gradual and abrupt genetic changes.

J. I. Gent · N. J. Nannas · R. Kelly Dawe
Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

Y. Liu · H. Su · F. Han
State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

H. Zhao · J. Jiang
Department of Plant Biology, Department of Horticulture, Michigan State University, East Lansin, MI 48824, USA

Z. Gao · J. A. Birchler (✉)
Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA
e-mail: BirchlerJ@Missouri.edu

5.1 Introduction

Centromeres are the sites on chromosomes where microtubule spindles attach during cell division. A set of proteins bind to centromeres during this process to constitute the large complexes called kinetochores that link centromeres to the spindle and regulate chromosomal dynamics. Other proteins, such as the centromeric histone H3

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_5

variant CENH3 (also known as CENP-A in many animals and fungi) bind to DNA. With the exceptions of broken chromosomes and B chromosomes, maize chromosomes are metacentric, as clearly seen by the primary constrictions in the center of metaphase sister chromatids (Fig. 5.1). Barbara McClintock pioneered research on maize centromeres in her creation of dicentric chromosomes in the 1930s, which resulted in two centromeres on a single chromosome pulling toward opposite poles and led to her discovery of the breakage-fusion-bridge cycle (McClintock 1939). Similar cytogenetics tools continue to be the source of discoveries about centromeres today, but the pace of discovery has increased dramatically with the advent of next-generation sequencing and centromere reference sequences.

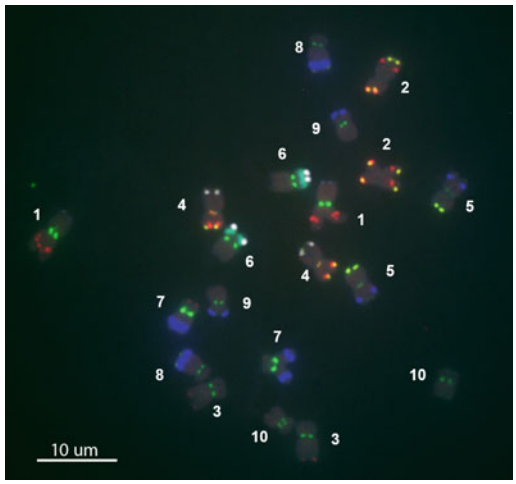

Fig. 5.1 Karyotype of inbred line B73. Chromosomal features of inbred line B73 are labeled by different fluorescent probes. The centromeric repeat, *CentC*, is green at the primary constriction of each chromosome. Note the variability in intensity among different chromosomes. Other repetitive sequences in green include the NOR on the short arm of chromosome 6 and a subtelomeric repeat (4-12-1). Features labeled in red include *Cent4* near the primary constriction of chromosome 4, the TAG microsatellite (1-26-2) arrays and another subtelomeric repeat (pMTY9ER). The 5S RNA cluster on the long arm of chromosome 2 is yellow. Features in blue are the 180 knob heterochromatin arrays. The white labels are for the *TR-1* type heterochromatin knobs. Preparation of root tips and chromosome painting was as described (Kato et al 2004). Bar = 10 um. Photo by Zhi Gao

5.2 Genetic Composition of Maize Centromeres

5.2.1 Centromere Repeats

Centromeres are regularly associated with satellite DNA (simple tandem repeats). However, the major satellite of maize, the 180 bp repeat called *knob180,* is exclusively on chromosome arms and visible as heterochromatic knobs (Fig. 5.1). A much less abundant satellite, the 156 bp *CentC*, is present in centromeres but in variable quantities (Albert et al. 2010) (Fig. 5.1). In fact, there is 20-fold more *knob180* than *CentC* in B73, and up to 200-fold more in other maize varieties (Bilinski et al. 2015; Gent et al. 2017). The centromeres with little *CentC* are composed of complex sets of genetic elements, mainly diverse retrotransposons, but also including transcriptionally active protein-coding genes (Zhao et al. 2016). Among the retrotransposons in maize centromeres are some that are enriched in centromeres and thus named centromeric retrotransposons (CRMs) (Nuemann et al. 2011; Presting et al. 1998; Miller et al. 1998; Sharma and Presting 2014). These CRMs specifically accumulate in centromeric chromatin, presumably due to chromodomains and putative centromere targeting by their integrases.

5.2.2 Simple and Complex Centromeres

The low abundance of satellites and complexity of other genetic elements in some maize centromeres provides an experimental resource for studying centromeres because the sequence complexity allows for assembly into the genome reference sequence and because it allows for unique mapping of sequence reads within centromeres (Wolfgruber et al. 2009, 2016; Jiao et al. 2017). We refer to such centromeres as complex centromeres. The maize reference inbred, B73, has seven complex centromeres and three dominated by *CentC* arrays; however, the number and location of complex centromeres varies widely among maize varieties and in wild

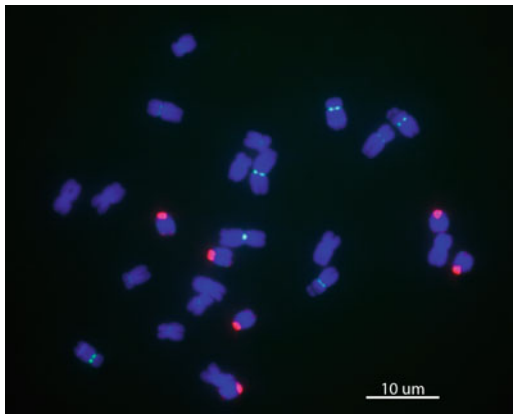

Fig. 5.2 Karotype of inbred line B73 plus B chromosomes. The chromosomes are stained with DAPI (blue). *CentC* (green) labels the centromeres. The B chromosomes are distinguished by the presence of the B-specific repeat (red) mainly in and around the centromere but also with a representation at the tip of the long arm of the B. Note also the presence of *CentC* along the length of the B chromosome. Bar = 10 um. Photo by Zhi Gao

relatives (Albert et al. 2010; Gent et al. 2015, 2017; Schneider et al. 2016; Zhao et al. 2017). Centromeres at unequal positions or with dissimilar sequence composition on homologous chromosomes in hybrids do not appear to perturb meiotic chromosome pairing, and the centromeres do not shift their positions on chromosomes to equilibrate (Lamb et al. 2007; Mondin et al. 2014; Gent et al. 2017). The supernumerary B chromosome has a specific repeat present in and around its centromere in addition to the elements common to the A centromeres (Lamb et al. 2005; Jin et al. 2005) (Fig. 5.2).

5.2.3 Centromere Positioning: Stable but Responsive to Genetic Change

The fact that maize centromeres are composed of such diverse genetic elements illustrates an important feature of centromeres: They are defined by their chromatin rather than their sequence, and in that sense, centromeres are thought of as epigenetic structures. Specifically, they are operationally defined by the presence of the H3 variant CENH3, a convention we use in this review. CENH3 occupies a subset of nucleosomes in centromeres, and chromatin immunoprecipation of CENH3 followed by sequencing (ChIP-seq) provides precise locations of CENH3 nucleosomes and hence centromeres. Another strong indication of the epigenetic nature of centromeres is the *de novo* formation of centromeres (neocentromeres) on broken chromosome arms (Fu et al. 2013; Zhang et al. 2013a, b, c; Liu et al. 2015). Related to this, maize centromeres expand when maize chromosomes are transferred into oat, which indicates that the DNA sequences on the chromosomes do not strictly define the size of centromeres. Nonetheless, all heritable variation in maize centromere positions observed to date is associated with major changes in the centromere DNA sequence or chromosome structure. Minor shifting in centromere position does occur at the purely epigenetic level, as evidenced by CENH3 ChIP-seq of individual plants, but it is not heritable and does not accumulate into large shifts over generations (Gent et al. 2017; Wang and Dawe 2018). In the absence of genetic change, centromere positions are usually stably maintained (Gent et al. 2015).

5.2.4 Influence of Genetic Elements on Centromere Positioning

The stability of centromere positioning suggests that genetic elements in or flanking centromeres might contribute to centromere positioning. A strong candidate for reinforcing centromere stability is euchromatic, transcribed genes (Wang et al. 2014), which are not conducive to CENH3 accumulation. Heterochromatic repetitive DNA such as *CentC* and retrotransposons, in contrast, likely positively reinforce centromere positioning. Even within centromeres, CENH3 occupancy is higher at such repetitive elements than at genes (Gent et al. 2015; Zhao et al. 2016). *CentC* is of particular interest because of its abundance, especially in wild *Zea* species and in *Tripsacum* (Bilinski et al. 2015; Gent et al. 2017). Surprisingly, however, despite conservation of its

consensus sequence across species, there appears to be little constraint on its sequence within individual genomes. Instead, individual copies are highly polymorphic and show no evidence for homogenization even within species with abundant *CentC*. *CentC* is also not confined to active centromeres (Lamb et al. 2005; Gao et al. 2011; Wang et al. 2014; Bilinski et al. 2015). It is not clear whether complex centromeres arise as a consequence of *CentC* deletion or whether formation of complex centromeres makes non-centromeric *CentC* susceptible to deletion. In most cases, even complex centromeres have some *CentC* within or very close to them (Albert et al. 2010; Gent et al. 2017). It is also not clear whether the ancestral state of *Zea* and *Tripsacum* represents complex or repetitive array centromeres. At present, *CentC* dominates centromeres in most *Zea* genomes and presumably in *Tripsacum dactyloides*, with its abundant *CentC* (Lamb and Birchler 2006) but the related genera, *Sorghum* and *Miscanthus,* lack *CentC* (Melters et al. 2013). It is clear, however, that the presence of large arrays of *CentC* usually indicates the locations of active centromeres, which suggests compatibility with centromere function. Any such contribution would not be the consequence of the specific linear sequence of *CentC*, but of some other property, for example, facilitating a specific form of heterochromatin (Gent et al. 2014; Zhao et al. 2016; Su et al. 2017), non-coding transcription (Gent and Dawe 2012), nucleosome positioning (Gent et al. 2011; Zhao et al. 2016), and/or DNA repair (Osman et al. 2013; Wolfgruber et al. 2016).

5.3 Centromere Chromatin Structure and Transcription

5.3.1 Features of CENH3 Nucleosomes

Maize centromeric chromatin is clearly distinguished from flanking pericentromeric chromatin in interphase cells by the presence of CENH3 and at least one other centromeric chromatin protein, CENP-C (Dawe et al. 1999). The amount of CENH3 present in maize centromeres is unknown, and it is not clear what affect CENH3 has on nucleosome or on higher-order chromatin structure. Digestion of chromatin by micrococcal nuclease produces similar sizes of DNA fragments from CENH3 nucleosomes as from total nucleosomes, approximately 150 bp, under moderate digestion conditions, but produces smaller fragments from CENH3 nucleosomes under more severe digestion conditions (Gent et al. 2011; Zhao et al. 2016). Similar observations have been made with micrococcal nuclease digestion of human CENP-A chromatin, which is attributed to weaker interactions between the nucleosome and DNA at its entry and exit points (Hasson et al. 2013). The spacing of CENH3 nucleosomes, approximately 190 bp between start positions, is indistinguishable from total nucleosomes (Gent et al. 2011; Zhao et al. 2016). Nucleosomes tend to reproducibly occupy the same positions relative to *CentC* and *CRMs*. In *CentC*, these positions tend to occur at approximately 10 bp intervals, corresponding to AA/TT dinucleotides (Gent et al. 2011), a phenomenon also observed in rice CENH3 nucleosomes (Zhang et al. 2013a, b, c). Several specific positions are strongly favored for nucleosomes on *CentC* (Zhao et al. 2016). 10 bp spacing of AA/TT dinucleotides can produce strong nucleosome/DNA interactions (Trifonov and Nibhani 2015), which could be advantageous in light of tension on the chromatin during mitosis.

5.3.2 Other Centromeric Chromatin Modifications

The timing of centromere replication, mid S phase rather than late S phase, distinguishes it from pericentromeric heterochromatin (Wear et al. 2017). Immunofluorescence experiments with 5-methylcytosine antibody revealed a striking lack of signal in centromeres relative to pericentromeres (Zhang et al. 2008; Koo et al. 2011). Bisulfite sequencing, however, has revealed little reproducible difference between centromeres and pericentromeres (Gent et al. 2012) and indicated that centromeres have highly

methylated DNA that is characteristic of heterochromatin (Gent et al. 2014; Su et al. 2016). Maize centromeres are deficient in RNA-directed DNA methylation and its associated siRNAs, but so too are pericentromeres and most other heterochromatin. A theoretical explanation for the discrepancy between immunostaining and sequencing results could be that centromere/kinetochore proteins obstruct access of the DNA to antibody binding under the immunofluorescence conditions that are otherwise suitable for the remainder of the genome. Phosphorylation of histone H2A at threonine 133 (H2AThr133ph) marks centromeres even in interphase (Dong and Han 2012; Su et al. 2017), and phosphorylation of histone H3 at threonine 3 (H3T3ph) marks centromeres during cell division (Liu et al. 2017). Other than H2AThr133ph, no examined histone modifications clearly differentiate the centromere from other heterochromatin in interphase cells. The heterochromatic H3 modifications dimethylation of lysine 9 and lysine 27 (H3K9me2 and H3K27me2) are relatively reduced in centromeres, which may be a consequence of reduced canonical H3 because of replacement by CENH3 (Gent et al. 2014; Zhao et al. 2016). None of the four euchromatic histone modifications examined, H3K4me3, H3K36me3, H3K27me3, nor H3K9ac have significant levels in centromeres or pericentromeres except in genes (Gent et al. 2012; Zhao et al. 2016). Meiotic crossover frequencies in centromeres are also low, as in other heterochromatin (Gore et al. 2009; Shi et al. 2010).

5.3.3 Non-coding Transcription of Centromeres

RNA from maize centromeric elements has been reported (Topp et al. 2004). Northern blots probed with DNA probes for *CentC* and a specific CRM named *CRM2* to detect RNA co-immunoprecipated with CENH3 produced signal corresponding to a range of RNA sizes (Topp et al. 2004). CENP-C is capable of binding to RNA, which can facilitate its binding to DNA in vitro (Du et al. 2010). RNA fluorescence in situ hybridization (RNA-FISH) with probes for *CentC* produced signal in nucleoli (Koo et al. 2016). Neither CENH3 nor CENP-C was detectable in nucleoli by immunofluorescence, but the kinetochore component MIS12 was, and its nucleolar localization was lost upon treatment with RNase (Koo et al. 2016). Some level of transcription of CRMs, at least of their polyprotein genes, is required for their maintenance in the genome. Full-length CRMs encode their own transcriptional units driven by RNA polymerase II (Neumann et al. 2011). Transcription of *CentC* is more enigmatic. Integration of CRMs or other transposons into *CentC* arrays could drive its transcription. Hypothetically, either the transcripts themselves could have a function, e.g., in facilitating assembly of the kinetochore, or the act of transcription itself could by facilitating CENH3 incorporation into nucleosomes.

5.4 Centromere and Kinetochore Proteins

5.4.1 Experimentally Confirmed Proteins in Maize

The kinetochore is comprised of many proteins; there are over 50 identified in yeast (Biggins 2013) and over 80 identified in humans (Cheeseman and Desai 2008). The structure and function of many of these proteins are highly conserved, which has allowed their identification in a range of organisms (Meraldi et al. 2006; Przewloka and Glover 2009). Four structural kinetochore proteins have been identified in maize: CENH3 (Zhong et al. 2002), CENP-C (Dawe et al. 1999), NDC80 (Du and Dawe 2007) and MIS12 (Li and Dawe 2009). CENH3 and CENP-C are constitutive components of the inner kinetochore that interact with centromeres, while NDC80 and MIS12 are members of the outer kinetochore that promote interactions with microtubules during cell division (Cheeseman and Desai, 2008). Three additional kinetochore proteins have been identified that monitor

attachments to the spindle and ensure correct segregation of chromosomes: MAD2 (Yu et al. 1999), BUB1, and BUB3 (Su et al. 2017).

5.4.2 CENH3

CENH3 is the foundation of the kinetochore in most species and helps recruit other kinetochore proteins such as CENP-C (Kato et al. 2013). Maize CENH3 was identified based on homology to its H3 counterpart, with 56% identity over the entire sequence and 78% identity within the histone-fold domain (Zhong et al. 2002). Maize CENH3 is a 17 kDa protein, 157 amino acids in length, and produced by a single-copy gene, *CenH3* (Zhong et al. 2002). As in other organisms, the N-terminal tail of maize CENH3 is longer than the tail of histone H3 and is diverged greatly in sequence (Henikoff et al. 2000; Talbert et al. 2002). Divergence in CENH3 N-terminal tails is so extreme that it is not possible to align sequences across taxa (Malik and Henikoff 2003; Maheshwari et al. 2015). The loop-1 domain of CENH3 is also highly divergent; this region contacts DNA, and its longer sequence compared to H3 has been proposed to confer greater sequence specificity (Malik and Henikoff 2003). Sequence specificity of CENH3, however, must be limited for its own positioning given the predominance of complex maize centromeres and the ability of maize CENH3 to reproduce the same DNA localization patterns as native CENH3 when transformed into Arabidopsis (Maheshwari et al. 2017).

Maize CENH3 localizes to the inner kinetochore of all chromosomes in all stages of the cell cycle in both mitotic and meiotic cells (Zhong et al. 2002). It co-localizes with another known maize inner kinetochore protein, CENP-C. While direct interaction between maize CENH3 and CENP-C has not been demonstrated, human versions have been shown to interact (Carroll et al. 2010). A CENH3-YFP fusion has been stably transformed into maize, and its incorporation into kinetochores has allowed visualization of active centromeres (Jin et al. 2008). The fusion has been used to demonstrate the interspersion of CENH3 and H3 on centromeres (Jin et al. 2008), identify phosphorylation patterns within the centromere (Dong and Han 2012), and study the mechanisms of haploid induction (Zhao et al. 2013). As in Arabidopsis, perturbing the structure of CENH3 in maize can induce haploidy, though at lower frequencies (Kelliher et al. 2016; Ravi and Chan 2010).

5.4.3 CENP-C

CENP-C is another conserved constitutive member of the inner kinetochore, and the maize homolog was identified based on a conserved 23 amino acid sequence known as region I (Brown 1995; Dawe et al. 1999; Meluh and Koshland 1995). Maize has three variants of CENP-C: CENPCA, CENPCB, and CENPCC. CENPCB is produced by a single-copy gene, while CENPCA and CENPCC are produced by multi-copy genes. *CenpcA* and *CenpcC* are 99.9% identical but diverge in the C-terminal coding region and 3' untranslated region. The resulting CENPCA and CENPCC proteins share 95% identity, while CENPCB shares 76–78% identity with A and C (Dawe et al. 1999).

Using both fluorescence tagging and immunolocalization, CENP-C has been shown to localize to the inner kinetochore and associate closely with CENH3 and the centromere (Dawe et al. 1999; Du et al. 2010). CENP-C is present at kinetochores throughout the cell cycle, but accumulates in interphase between G1 and G2 (Dawe et al. 1999). CENP-C has DNA-binding capabilities, and it has been shown to directly bind the *CentC* centromere repeat in vitro (Du et al. 2010). However, this interaction is non-sequence specific in these assays because CENP-C can be competed away from *CentC* with non-centromere sequences. CENP-C also binds RNA non-specifically using the same C-terminal DNA-binding subdomain located in exons 9–12. This interaction does not compete with DNA. In fact, addition of single-stranded RNA promotes and stabilizes binding of CENP-C to DNA in vitro. Using stably transformed YFP-CENP-C fusions, it has been shown

that deletion or substitution of the exon 9–12 subdomain reduces kinetochore localization by 20% (Du et al. 2010). Additionally, tagging the C-terminus of CENP-C also disrupts kinetochore localization (Du et al. 2010).

5.4.4 NDC80

NDC80 is a broadly conserved outer kinetochore protein that links inner kinetochore proteins with microtubules (Varma and Salmon 2012). It directly attaches to microtubules through a positively charged tail and calponin homology domain found in the N-terminus (DeLuca and Musacchio 2012). The maize NDC80 homolog was identified through these conserved regions (Du and Dawe 2007). It shares more than 40% similarity with yeast, *Xenopus*, chicken and human versions of NDC80. Maize NDC80 is a 576 amino acid protein with a molecular mass of 75 kDa produced from a single-copy gene (Du and Dawe 2007). While microtubule-binding capacity has not been demonstrated, maize NDC80 does localize to the outer kinetochore. Using a peptide antibody to the N-terminus, it was shown that maize NDC80 localizes beyond CENP-C and CENH3 on the outer face of the kinetochore where microtubules associate (Du and Dawe 2007), consistent with NDC80 in animals (DeLuca et al. 2006). Surprisingly, however, maize NDC80 is also constitutively present on kinetochores throughout the cell cycle, unlike other species where it is present only during cell division (Hori et al. 2003). NDC80 localizes with another maize kinetochore protein, MIS12, and helps form a bridge that holds sister kinetochores together in meiosis I, as discussed further below.

5.4.5 MIS12

MIS12 is part of a complex that interacts directly with the NDC80 complex and indirectly with CENH3 and CENP-C (Cheeseman and Desai 2008). It is required to maintain the structural integrity of the kinetochore (Kline et al. 2006). Maize has two copies of the *Mis12* gene, *Mis12-1,* and *Mis12-2*, which share 89% sequence identity but produce two different sized proteins. MIS12-2 is a 244 amino acid protein, while MIS12-1 is a 223 amino acid protein due to an early stop codon (Li and Dawe 2009). Both genes are expressed, but with *Mis12-2* at a higher level. As with CENH3, CENP-C, and NDC80, MIS12 proteins localize constitutively to the kinetochore throughout the cell cycle. In meiosis, homologous chromosomes separate in meiosis I, and sister chromatids separate in meiosis II. MIS12 co-localizes with NDC80 on the outer kinetochore, and in meiosis I, they form the "NDC80-MIS12 bridge" that holds sister kinetochores together to help prevent premature separation of chromatids (Li and Dawe 2009). RNAi knockdown of MIS12-1 and MIS12-2 causes 30% of sister kinetochores to separate prematurely in meiosis I, producing failures in chromosome alignment, uneven microtubule bundles, and stalled chromosomes in anaphase (Li and Dawe 2009). Anaphase I disjunction failures cause errors in meiosis II spindle assembly and tetrad formation with isolated chromosomes forming independent mini-nuclei (Li and Dawe 2009).

5.4.6 MAD2, BUB1, and BUB2

Chromosomes must attach to the spindle in the proper orientation to ensure that each daughter cells receives one copy. The spindle assembly checkpoint monitors and breaks down incorrect attachments between kinetochores and microtubules. MAD2 is one of the major components of the spindle checkpoint; this highly conserved protein localizes on kinetochores that are improperly attached to the spindle and delays cell cycle progression until all attachments are correct (Musacchio and Salmon 2007). The single-copy maize homolog of MAD2 was identified based on similarity to MAD2 homologs of human, Xenopus and yeast (64–70% similarity) (Yu et al. 1999). *Mad2* produces a 208 amino acid protein with a molecular mass of 24 kDa. In maize mitosis, MAD2 is absent from kinetochores in

prophase but appears in prometaphase. Its localization on the outer kinetochore is correlated with microtubule attachment; once microtubules associate with the kinetochore, MAD2 staining disappears. Treating cells with microtubule depolymerizing drugs causes intense localization of MAD2 on kinetochores, demonstrating that the maize spindle checkpoint is activated by unattached kinetochores. MAD2 remains on kinetochores until the chromosomes have aligned on the metaphase plate and are under sufficient tension (Yu et al. 1999).

BUB1 and BUB3 are also components of the spindle checkpoint that localize on kinetochores (Lara-Gonzalez et al. 2012). In humans, BUB1 and BUB3 localize to the outer kinetochore and serve as a scaffold for coordinating other checkpoint proteins (Elowe 2011). BUB3 is crucial for stabilizing kinetochore–microtubule attachments (Logarinho et al. 2008). Unlike MAD2, localization of BUB1 and 3 on kinetochores is not an indicator that the spindle checkpoint is active, rather they make the kinetochore competent to signal the checkpoint should it be necessary (Lara-Gonzalez et al. 2012). Maize homologs of BUB1 and BUB3 have been identified (Su et al. 2017). BUB1 is a 553 amino acid protein with a highly conserved Ser/Thr kinase domain (amino acids 249–543). It lacks the GLEBS domain identified in other species that promotes interaction with BUB3 and other checkpoint proteins, suggesting its function may have diverged (Elowe 2011). Maize BUB3 is a 343 amino acid protein with seven WD40 domains that have been shown to promote protein–protein interactions in other species (Smith et al. 1999; Su et al. 2017). Both proteins localize to mitotic kinetochores in maize beginning in interphase and increase through prophase, but show reduced localization in metaphase and anaphase. In meiosis, BUB1 and BUB3 localize weakly in interphase and peak in pachytene of prophase I. Both proteins remain at the outer kinetochore in metaphase and anaphase I and throughout all stages of meiosis II (Su et al. 2017). BUB1 is also involved in phosphorylation of histone H2A; RNAi knockdown of maize BUB1 caused a reduction in H2AThr133 phosphorylation levels in certain tissues (Su et al. 2017).

5.5 Maize Centromeres in Oat

5.5.1 Use of Oat–Maize Addition Lines to Study Centromere Structure

Because most of the centromeric repeats can be located in both centromeric and pericentromeric regions as well as in most or all centromeres in the same species, it is technically challenging to study the fine structure of a specific centromere in most plant species. Oat–maize chromosome addition (OMA) lines provide unique materials to study individual maize centromeres. The OMA lines were developed from crosses between oat and maize (Kynast et al. 2001; Rines et al. 2009). In such crosses, the maize chromosomes are usually lost to create an oat haploid. However, maize chromosomes can become established and perpetuated in oat at low frequency. Chromosome doubling of these events generates disomic OMA lines that contain all 42 oat chromosomes and a pair of maize chromosomes, which are relatively stable. Thus, disomic OMA lines can be readily maintained and used for various cytological and molecular studies of individual maize chromosomes present in oat (Fig. 5.3).

The *CentC* satellite repeat is in most maize centromeres but is absent from oat centromeres (Jin et al. 2004). Thus, it provides an excellent DNA mark to track individual maize centromeres in oat. The structure and sizes of the *CentC* repeat arrays in each maize centromere can be analyzed via fiber-FISH mapping of the OMA lines (Jin et al. 2004). Maize centromeres contain variable amounts of intermingled *CentC*-CRM arrays, ranging from ~300 Kb to several megabases based on such fiber-FISH measurements. The association of *CentC* and CRM with CENH3 can be directly visualized by immunofluorescence on stretched maize centromeres in OMA lines (Jin et al. 2004). The cytological studies also revealed that not all

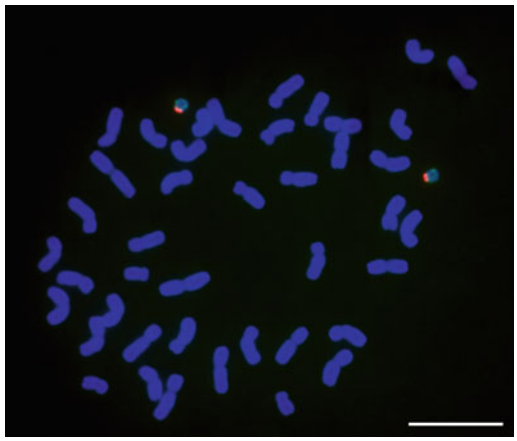

Fig. 5.3 B chromosomes in an oat background. The chromosomes are stained with DAPI (blue). The maize B chromosome is labeled with the B-specific repeat (red) and *CentC* (green). Note that *CentC* is not present in oat chromosomes. Bar = 10 um

CentC or CRM sequences are associated with CENH3 (Jin et al. 2004).

5.5.2 Expansion of Maize Centromeres in Oat

The increasingly improved throughput of next-generation sequencing techniques allowed sequencing-based analyses of the maize centromeres in oat. Although >95% of the sequence data obtained from OMA lines are derived from the oat genome, the amount of maize sequences from the OMA lines are sufficient for genomic and epigenomic studies of individual maize centromeres (Wang et al. 2014; Zhao et al. 2016). In addition, maize and oat sequences are sufficiently diverged to allow unequivocal identification of the maize sequences. Chromatin immunoprecipitation using a CENH3 antibody followed by sequencing (ChIP-seq) in OMA lines resulted in a surprising discovery: the CENH3-binding domains of maize centromeres are dramatically expanded to encompass a larger area in the oat background (~3.6 Mb) than the average centromere size in maize (~1.8 Mb) (Wang et al. 2014). In addition, the direction of maize centromere expansion appears to be restricted by the transcription of genes located in regions flanking the original centromeres (Wang et al. 2014). Studies of maize centromeres in oat have revealed several fundamental features associated with centromere function: (1) The size of the centromere is regulated. Centromeres appear to maintain a uniform size within a species regardless of chromosome size. (2) The plasticity of centromeres is epitomized by the expansion of maize centromeres in oat. The expanded array of CENH3 nucleosomes can adapt to various types of DNA sequences located in the pericentromeric regions. (3) The expanded centromeric regions prefer non-transcribed sequences.

5.6 Centromere Evolution

5.6.1 Rearrangements of Maize Chromosomes Relative to Sorghum

Maize is believed to have originated from an ancient tetraploid that contained 40 chromosomes and underwent dramatic intra- and inter-chromosomal rearrangements to produce the current diploid 20 chromosomes (Whitkus et al. 1992; Paterson et al. 2004; Wei et al. 2007). Thus, most maize chromosomes are fused from two or multiple chromosomes from the progenitor tetraploid ancestor (Wei et al. 2007). Half of the 40 centromeres were either deleted or inactivated during the chromosome number reduction. The sorghum chromosomes are believed to maintain synteny with ancient chromosomes that predate the divergence of maize and sorghum (Wei et al. 2007; Schnable et al. 2009). Thus, comparative sequence analysis was used to determine which of the 20 ancient centromeres survived during maize genome evolution (Wang and Bennetzen 2012). For example, maize chromosome 3 was derived from two ancient chromosomes, which are homologous to sorghum chromosomes 3 and 8, respectively (Wang and Bennetzen 2012). An ancient chromosome 8 inserted in the pericentromeric region of an ancient chromosome 3 and the fused

chromosome underwent several intrachromosomal rearrangements that resulted in the current maize chromosome 3 (Wang and Bennetzen 2012; Zhao et al. 2017). The centromere of maize chromosome 3 (*Cen3*) was derived from ancient chromosome 3 and the centromere of ancient chromosome 8 became inactivated after the fusion of the two chromosomes (Zhao et al. 2017).

5.6.2 CentC

As noted above, *CentC* is a 156-bp satellite repeat that is enriched in centromeres (Ananiev et al. 1998). Most maize centromeres contain intermingled *CentC*-CRM arrays, ranging from none to several megabases (Jin et al. 2004). The *CentC* repeat shows substantial sequence similarity with the centromeric satellite repeats from several distantly related grass species, including the 155-bp *CentO* repeat from rice, but excluding some closely related species such as sorghum (Cheng et al. 2002; Lee et al. 2005; Melters et al. 2013). A conserved 80-bp domain was found among the centromeric repeats from various grass species (Lee et al. 2005; Yang et al. 2018), suggesting an evolutionary constraint. The *CentC* repeats show high sequence similarity with minor sequence differences in the form of point mutations (Ananiev et al. 1998). The average identity between *CentC* copies to the consensus is about 96% (Gent et al. 2017). However, analyses of long stretches of *CentC* arrays associated with centromere 2 and centromere 5 showed that the *CentC* repeats are more homogenized within clusters, supporting the notion that tandem repeats can increase copy number by local duplication (Bilinski et al. 2015). Interestingly, with the exception of the *huehuetenangensis* sub-species, the *CentC* repeats are significantly more abundant and more uniformly present in all centromeres in teosintes, the close wild relatives of maize (Albert et al. 2010). Thus, reduced amounts of *CentC* occurred within the timeframe of maize domestication (Schneider et al. 2016).

5.6.3 Evidence for Ancient *de Novo* Centromeres

Centromeres can be activated *de novo* in genomic regions devoid of any centromere-specific sequences, and these are often called "neocentromeres" in other species. Neocentromeres were first discovered in humans and have been reported in several plant species (Nasuda et al. 2005; Gong et al. 2009; Topp et al. 2009; Fu et al. 2013; Liu et al. 2015). A chromosomal rearrangement that deleted a portion of the native centromere of a maize chromosome in the genetic background of oat resulted in a new centromere position that was shifted over a megabase relative to the original chromosome (Topp et al. 2009; Wang et al. 2014), and maize chromosomal rearrangements that completely lack native centromeres produce novel centromere positions (Fu et al. 2013; Liu et al. 2015) as described in more detail below. Remarkably, the centromere position for a specific maize chromosome is not fixed among different maize inbred lines, possibly indicative of ancient *de novo* centromere formation. For example, three major centromere positions were observed on maize chromosome 5, including CEN5L located at 102.1–103.7 Mb, CEN5 M at 105.2–106.8 Mb, and CEN5R at 107.9–109.8 Mb with distinctions in the CENH3 association pattern among these (Schneider et al. 2016). Multiple centromere positions were observed for most chromosomes in different maize lines and in wild relatives (Schneider et al. 2016; Zhao et al 2017). Dating CRM elements within centromeres raised the possibility that the repositioning occurred after maize domestication (Schneider et al. 2016). It was proposed that strong selection for centromere-linked genes in domesticated maize may have reduced the

diversity of the maize centromeres to only one or two post-domestication haplotypes (Schneider et al. 2016).

5.7 Classical Neocentromeres on Knobs and Meiotic Drive

While new stable centromeres marked by CENH3 are commonly called neocentromeres, the term was originally coined by Marcus Rhoades to describe ectopic formation of centromere-like activity on knobs in the presence of abnormal chromosome 10 (Ab10) (Rhoades and Vilkomerson 1942). The clearly visible movement of knobs at anaphase I and II in Ab10 lines is associated with a preferential segregation, or meiotic drive, of knobs and linked loci through female meiosis (Rhoades 1942, 1952). Further studies of Ab10-mediated neocentromeres have revealed they are not controlled by the same proteins as those involved for true centromeres [not CENH3, CENPC, MIS12, NDC80, or MAD2 (Dawe and Hiatt 2004)]. At least two loci are involved in neocentromere activity, one that controls the movement of knobs containing 180 bp *(knob180)* repeats (shown in Fig. 5.1) and another that controls the movement of *TR-1* repeats (shown in Fig. 5.1) (Hiatt et al. 2002).

Genetic analysis has shown that the locus controlling 180 bp neocentromere activity is required for meiotic drive but that *TR-1* neocentromere activity is not (Kanizay et al. 2013). Both activities have been postulated to be driven by kinesin microtubule-based motors (Hiatt et al. 2002). Neocentromere activity of *knob180* repeats is now known to be caused by not one kinesin but a cluster of nine kinesin genes called *Kinesin driver*, or *Kindr* genes, located on Ab10 (Dawe et al. 2018). Stable epimutants that abolish expression of the *Kindr* complex also result in a loss of meiotic drive (Dawe et al. 2018), confirming the key role of neocentromere activity in meiotic drive as originally proposed by Rhoades (1952).

5.8 B Chromosome Centromeres, Nondisjunction, and Centromere Inactivation

5.8.1 Behavior and Structure of B Chromosomes

The B chromosome of maize is a supernumerary chromosome (Fig. 5.2) that is not vital nor is it detrimental unless in high copy number. Its dispensable nature is counteracted by an accumulation mechanism that maintains it in populations. This mechanism consists of nondisjunction at the second pollen mitosis followed by preferential fertilization of the egg by the B chromosome containing sperm (Roman 1947, 1948). These two properties keep the chromosome from drifting to extinction. The centromere in this near telocentric chromosome is the site of nondisjunction but other sites on the B chromosome are also needed for this process. Notably, the very tip of the long arm is required (Ward 1973). When it is removed from the same cell as the centromere, the latter will no longer undergo nondisjunction.

An experiment to find sequences found only on the B chromosome was successful in isolating the B chromosome-specific repeat, ZmBs (Alfenito and Birchler 1993). The unit length is a little more than 1.0 Kb, but the numerous copies analyzed are highly variable. Its sequence has similarities to the knob heterochromatin repeat and to telomere repeats. A similar sequence is present near the centromere of chromosome 4 (Page et al. 2001) in all maize lines examined (Albert et al. 2010). This repetitive B-specific sequence is scattered throughout and around the centromere of the B chromosome (Lamb et al. 2005; Jin et al. 2005). At the core of the B centromere are *CentC* and CRM sequences spanning about a 750 kilobase region (Jin et al. 2005). Minichromosomes containing basically just the centromere have *CentC* and CRM and the B-specific repeat (Kato et al. 2005; Han et al. 2007a, b). These minichromosomes can be

induced to undergo nondisjunction in the presence of a full-sized B chromosome that provides the needed trans-acting factors in the long arm (Masonbrink and Birchler 2012). This result raises the possibility that the B-specific repeat is the target for nondisjunction because it is unique to the B chromosome centromere while the other centromeric repeat elements present are also located at A centromeres that lack nondisjunction.

5.8.2 Misdivision Analysis of B Centromeres

Misdivision involves the centromere being attached to both poles in metaphase and then ripped apart during anaphase. With regard to the maize B chromosome centromere, Carlson found several misdivision derivatives (Carlson 1973; Carlson and Chou 1981). He used a B-A translocation involving the short arm of chromosome 9 (TB-9Sb). This arm carries several excellent endosperm markers that facilitate its manipulation. Carlson recognized that misdivisions of this centromere in the previous meiosis could be recognized in the progeny kernels because they exhibit a mosaicism for the 9S markers in the endosperm. This is likely related to the breakage-fusion-bridge (B-F-B) cycle, although the mechanics with regard to the centromere are unclear (Kaszas and Birchler 1998). However, just as with the B-F-B cycle, the broken chromosomes are "healed" in the sporophytic embryo so the broken centromere or its fusion onto itself as an isochromosome is captured in its initial form in the specific embryo associated with the selected mosaic endosperm.

This phenotypic screen was used to isolate a large collection of misdivision events for molecular analysis of the B chromosome centromere (Kaszas and Birchler 1996). Using this approach in successive misdivisions, the size of the B centromere could be reduced progressively as followed by the complexity of the B-specific repeat in Southern blots. The main core of the B centromere was subsequently estimated from a selection of these derivatives as being approximately 750 Kb (Jin et al. 2005). The smallest centromeres recovered were in the range of a few hundred Kb (Kaszas and Birchler 1998). Whether this size is an indication of the minimal size of a functional centromere is still open to question given that it is now known that the domain of CENH3 can expand relatively quickly, for example, when introduced into oat, as described above.

5.8.3 Centromere Inactivation

The first recognized cases of centromere inactivation in plants were found in maize (Han et al. 2006). In the process of studying the chromosome type of breakage-fusion-bridge (B-F-B) cycle, structurally dicentric chromosomes were found to be stable. The B-F-B cycle was initiated using a foldback duplication of the short arm of chromosome 9, first generated by Barbara McClintock, that had been recombined onto TB-9Sb by Zheng and colleagues (Zheng et al. 1999) to produce TB-9Sb-Dp9. The advantage of this arrangement is that the B-F-B cycle can continue throughout the life cycle because it is operating on a dispensable chromosome. The stable chromosomes recovered by Zheng and colleagues were minichromosomes with B chromosome centromeres that were nevertheless stable. Han et al. (2006) extended this analysis with the recovery of additional examples. Of unusual note were several such minichromosomes with apparently two sets of centromeres. When these were examined for the presence of CENH3, it was found that only one of the two sets of centromere sequences was associated with this centromeric histone, suggesting that the other set was inactive (Han et al. 2006). Indeed, examination of meiosis revealed that only the one associated with CENH3 progressed to the poles.

Another product of this experiment was the recovery of an inactive B centromere that is present on the tip of chromosome arm 9S (9-Bic-1) (Han et al. 2006) (Fig. 5.4). The process by which the broken chromosome became attached to the same arm from which it was derived is not known. Homozygotes for this

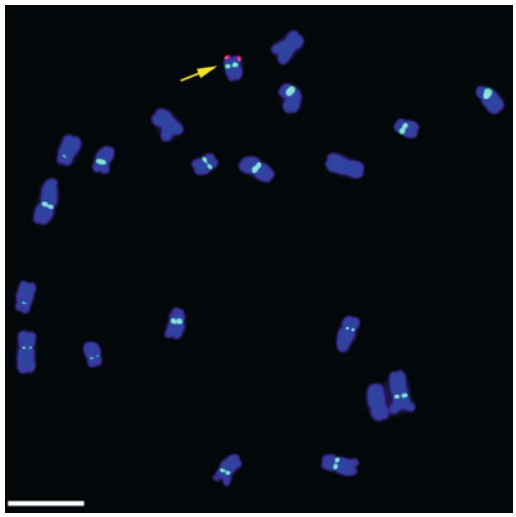

Fig. 5.4 Karyotype of a plant with 9-Bic-1 with an inactive B centromere at the tip of chromosome arm 9S. Chromosomes are stained with DAPI in blue. *CentC* is green and the B-specific centromere repeat is red. The arrow denotes chromosome 9 with the appended inactive B centromere. Bar = 10 um

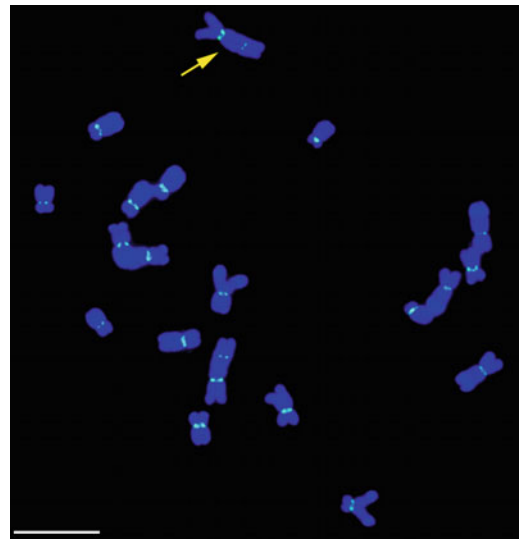

Fig. 5.5 Karyotype of Translocation 1–5 with an inactive centromere. Chromosomes are stained with DAPI in blue. Centromeres are labeled with a CRM probe (green). The arrow notes one of the T1-5 chromosomes. Note that only one of the CRM sites is present at a primary constriction; the other site is the inactive centromere. Bar = 10 um

chromosome have a chlorotic phenotype suggesting that some genes are missing from the tip of 9S. Nevertheless, this example has proven to be very useful for studies of inactive centromeres because it has stable transmission from one generation to the next. Fiber-FISH analysis indicates that the whole of the B centromere is present (Jin et al. 2005). There is no association with CENH3 (Han et al. 2006).

With the realization that centromeres could become inactive, other cases were soon found. A translocation between chromosomes 1 and 5 that has been used extensively in maize genetics was revealed to have an inactive centromere at the junction of the two chromosomes (Gao et al. 2011) (Fig. 5.5). This translocation was recovered from material exposed to an atomic bomb test on an atoll in the Pacific Ocean in 1948. Apparently, the inactive state of the centromere had been perpetuated over the decades. Yet another example involved the Tama Flint line. Chromosome 8 was found to have two spatially separate sites of canonical centromeric sequences (Lamb et al. 2007). However, only one site was associated with CENH3 and was the one that progressed to the poles in meiosis. Interestingly, the active site was displaced by an inversion that positioned the centromere about 20% of the chromosome arm from the usual position. Heterozygotes between this chromosome and a normal one failed to show any evidence of recombination between the two, so there was no negative fitness associated with this arrangement. Still further, the long arm of the B chromosome was found to contain sites of typical centromeric sequences but without any evidence of their ability to organize a kinetochore (Lamb et al. 2005). Studies of *de novo* centromere formation suggest that recovery of potential misdivision derivatives of *de novo* centromeres showed evidence of centromere inactivation as well, as described below (Liu et al. 2015).

Centromere inactivation could also be directed. The B-9-Dp9 chromosome was crossed to misdivision derivatives of TB-9Sb with reduced amounts of centromeric sequences. Because the duplicated chromosome has a reverse duplication, it can recombine with misdivision derivatives of

TB-9Sb described above. From the heterozygotes of chromosomes with normal/small centromeres, crossing over will produce a dicentric between the large and small centromeres (Han et al. 2009). At anaphase I of meiosis, a bridge will be formed and the recombinants will be destroyed by breakage. However, at some frequency, dicentrics were recovered containing the large and small centromere together at opposite ends presumably tied together by recombination. One example that was inherited has been studied in detail. The small centromere shows no evidence of centromeric activity and no detectable CENH3. Thus, in the tug of war set up between the large and small centromeres by the recombination event, the small centromere lost and became inactive. This chromosome could be maintained in this state over many generations.

The nature of this recovered chromosome is that it is a foldback structure with the large active centromere at one end and the small inactive centromere at the other. This foldback chromosome is capable of recombining with itself (Han et al. 2009). Some of the recombinant products will produce a dicentric of the large centromere, which forms a bridge in anaphase and is destroyed. The other potential product is a dicentric of the two small centromeres. Interestingly, these products have been recovered and there is now detectable CENH3 on the chromosome (Han et al. 2009). These cases might be examples of re-activation of an inactive centromere. However, with the realization of the high frequency of *de novo* centromere formation on otherwise acentric fragments in maize, it is possible that such an event could explain their regular recovery. Because there is as yet no reference sequence for the B chromosome to determine if there is re-association of CENH3 with the progenitor sequences, it is not possible to rule out this possibility.

5.8.4 *de Novo* Centromere Formation on Chromosomal Fragments

Multiple cases of *de novo* centromere formation in the past century have been documented in the context of deletion of the native centromere. For example, examination of a chromosome fragment called Duplication 3a that had been induced by UV irradiation of pollen by Stadler and Roman in the 1940s revealed a *de novo* centromere (Fu et al. 2013). The frequent somatic loss of this chromosome led to the idea that it was a ring. However, at least in its present state, it is a linear chromosome with telomeric sequences at both ends but has no detectable *CentC* nor CRM repeats. When ChIP-seq was performed using antibodies against CENH3, a novel region of association spanning 350 Kb was found in the long arm of chromosome 3 that encompassed several unique genes. Another *de novo* centromere spanning 723 Kb of the pericentromere of chromosome 9 was formed in the mini-chromosome sDic15 (small dicentric chromosome 15) (Zhang et al. 201a, b, c). Similarly, 288 Kb of DNA from the short arm of chromosome 9 gave rise to a *de novo* centromere on Derivative 3–3, produced by a translocation between a B chromosome and chromosome 9 and a series of misdivisions (Liu et al. 2015) (Fig. 5.6). In Derivative 3-3-11 (itself a derivative of Derivative 3–3), a *de novo* centromere formed that spanned 200 Kb on the short arm of chromosome 9, with the progenitor *de novo* centromere of Derivative 3–3 being inactivated (Liu et al. 2015).

The DNA composition of these *de novo* centromeres is different from the native, established ones (Su et al. 2016). *De novo* centromeres on Dp3a and Derivative 3–3 were formed in euchromatic regions with lower retrotransposon

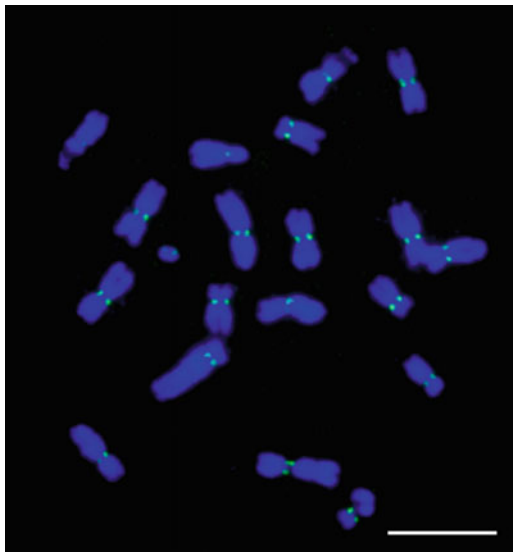

Fig. 5.6 CENH3 immunolabeling of derivative 3–3 with a *de novo* centromere. Chromosomes are labeled with DAPI in blue. The green signal indicates immuno-labeling with antibodies against CENH3. The small chromosome is derivative 3–3, which has no detectable *CentC* or CRM sequences but shows labeling with CENH3 on a *de novo* centromere. Bar = 10 um

density, while sDic15 has a de novo centromere in heterochromatin with high retrotransposon density (Su et al. 2016; Zhang et al. 2013a, b, c). These *de novo* centromeres indicate that CENH3 nucleosomes can load on DNA sequences with very different features. The DNA methylation level in the chromosomal regions before *de novo* centromere formation can be either high (in sDic15) or low (in Dp3a and 3–3) and attain the same level as native centromeres after *de novo* centromere formation (Su et al. 2016; Zhang et al. 2013a, b, c). Thus, the DNA methylation level does not appear to determine sites of *de novo* centromere formation but might be involved in centromere maintenance (Su et al. 2016). The transcription level of the progenitor sites is also related to the *de novo* centromere formation (Su et al. 2016). CENH3 nucleosomes prefer to load on regions with low or no expressed genes even on a fragment with high gene density (Su et al. 2016). The fact that chromosomal fragments that lost native centromeres were not immediately eliminated indicates that *de novo* centromeres organize on such fragments rather quickly, potentially within the span of a single cell cycle and at a maximum, only a few.

5.8.5 Competition Based on Centromere Size

If *de novo* centromeres can arise so quickly and apparently regularly, the question arises as to why chromosomes are not fragmented as dicentrics on a regular basis. A potential explanation worthy of test is suggested by the results from the centromere tug of war. In that case, if a small and large centromere were pitted against each other, the small centromere could become inactive, as is the case with a TB-9Sb derivative (Han et al. 2009). Thus, it is possible that *de novo* centromeres might arise in chromosome arms regularly but in opposition with the established centromere they are quickly epigenetically inactivated without a trace. Relative size difference between established centromeres and *de novo* ones would be the determining factor in such competitions and has been proposed to explain CENH3-mutant induced haploidy as well (Wang and Dawe 2018).

5.8.6 Centromere Activity not Required for B Chromosome Nondisjunction

The 9-Bic-1 dicentric chromosome was used to test whether the nondisjunction property of the B chromosome required centromeric activity (Han et al. 2007a, b). For this test, the 9-Bic-1 chromosome carrying an inactive B centromere was crossed to a line with normal B chromosomes. This configuration would supply the trans-acting factors needed, which are missing from the 9-Bic-1 chromosome because the distal B chromosome tip is not present. When this combination was used in an outcross as a male, the inactive B centromere caused chromosomal breakage at the second pollen mitosis in the presence of extra B chromosomes but was stable

otherwise. In some cases, the whole of chromosome 9 underwent nondisjunction as evidenced by two or zero copies being present in some progeny. The broken chromosome with the inactive B centromere could become attached to other chromosomes and in one instance was present at the terminus of chromosome 7. This case was homozygous viable, suggesting that the broken part of 9S was appended to the very tip of the short arm of 7. These results indicate that it is B centromere sequence rather than activity that is required for its nondisjunction property.

5.9 Centromere Pairing

Meiosis is a specialized eukaryotic cell division by which diploid cells undergo a single round of DNA replication and two rounds of chromosome segregation to yield haploid products. To achieve this result, parental homologs must pair during meiosis. The centromere has emerged as an important player in homologous chromosome pairing, and recent progress in a number of organisms suggests that centromere interactions in early meiotic prophase I are a general feature of meiosis (Da Ines et al. 2012; Kemp et al. 2004; Martinez-Perez et al. 1999; Phillips et al. 2012; Ronceret et al. 2009; Takeo et al. 2011; Tsubouchi and Roeder 2005; Unhavaithaya and Orr-Weaver 2013; Wen et al.. 2012; Zhang et al. 2013a, b, c). Centromere interactions include centromere clustering, centromere coupling, and centromere pairing. Centromere clustering refers to the situation in which centromeres are associated into groups. Centromere coupling is when there are non-homologous centromere associations, whereas centromere pairing refers to homologous centromere association.

In maize, centromeres do not associate in the pre-meiotic interphase; they begin to pair at leptotene and persist so until the pachytene stage (Zhang et al. 2013a, b, c). Centromere pairing occurs earlier than the telomere bouquet formation and pairing of chromosome arms but is important for initiation of homologous chromosome pairing (Zhang et al. 2013a, b, c). Unlike the non-homologous centromere clustering or coupling, centromere interactions of early meiotic prophase in maize are predominantly (about 65%) between homologous chromosomes. Centromere pairing depends on centromere activity; inactive centromeres cannot initiate this process even when homologous (Zhang et al. 2013a, b, c). In a maize line containing a stable dicentric chromosome, 7-Bic-1, a segment of a B chromosome containing an inactive centromere is translocated to maize chromosome 7 (Han et al. 2007a, b). At the leptotene stage, the inactive B centromere cannot pair as do the functional A centromeres. In another maize line containing a stable structurally dicentric reciprocal translocation between chromosomes 1 and 5, T1-5, there is an inactive A centromere at the translocation junction (Gao et al. 2011). At the leptotene stage, the inactive A centromere also does not pair. These results indicate that the centromeric DNA sequence is not a sufficient requirement for meiotic centromere pairing in maize, and homologous centromere interactions require functional centromeres.

Meiotic centromere pairing in maize is dependent on the presence of the REC8 cohesion protein (Zhang et al. 2013a, b, c). In the maize *afd1* mutant, which has a deletion of the maize homolog of the REC8 gene, centromeres do not pair at early prophase I. The process of centromere pairing in maize is independent of the synaptonemal complex (SC) central element protein ZYP1. ZYP1 is loaded onto the centromeric regions of chromosomes after centromere pairing. Another component of the central element of the maize SC, SMC6 (structural maintenance of chromosomes 6), is required for meiotic centromere pairing. The exact role of centromere interactions remains to be determined, but the available data suggest that centromere pairing leads homologous chromosome pre-alignment and facilitates the homology-scanning process.

5.10 Histone H2A Phosphorylation

In contrast to studies on phosphorylation of histone H3, including H3Ser10, H3Thr11, and H3Ser28 (Houben et al. 1999; Kaszas and Cande 2000; Zhang et al. 2005; Han et al. 2007a, b), little is known about histone H2A phosphorylation and its function in plants. Recently, the maize spindle assembly checkpoint (SAC) component Bub1-mediated phosphorylation of histone H2AThr133 was reported to localize in centromeric regions in plant mitotic and meiotic chromosomes during the entire cell cycle (Dong and Han 2012; Su et al. 2017). The immunostaining signals of H2AThr133ph in maize dicentric chromosomes reveal that only the functional centromere is phosphorylated, suggesting that H2AThr133ph is a new epigenetic marker for centromere function. Immunolocalization combined with ChIP-seq analysis revealed overlap between CENH3 and H2AThr133ph during interphase. H2AThr133ph signals spread to the pericentromeric and inner centromere regions during (pro)metaphase, but the strength of the signal drops during later anaphase and telophase. The presence and localization of H2AThr133ph was not changed in various maize lines showing precocious separation of sister centromeres, including minichromosomes, *afd1* mutants and *Mis12* RNAi transgenic lines, suggesting that H2AThr133ph is a stable feature of centromeres regardless of centromere orientation in meiosis I (Su et al. 2017). Histone phosphorylation is dynamically regulated during cell division, which coordinates with chromosome behavior during the cell cycle (Kouzarides 2007). The cell cycle-dependent H2AThr133 phosphorylation and the relationship with CENH3 nucleosomes may function temporally and spatially on the centromere morphology for proper chromosome segregation during cell division in plants.

5.11 Concluding Remarks

Classically, the centromere was thought to be inviolate and highly stable. However, in the past few years, it has been revealed that centromeres are variable for position in varieties of maize. Moreover, their activity can be silenced or *de novo* centromeres can arise with both states being perpetuated over generations. There is plasticity for the chromosomal domains over which they reside. The reduced crossing over around centromeres goes hand in hand with the variability of centromere position observed within a limited range. Thus, despite the remarkable malleability of centromeres, these constraints insure the continuity of the chromosome in development and over generations.

Acknowledgements Research supported by NSF grant IOS-1444514.

References

Albert PS, Gao Z, Danilova TV, Birchler JA (2010) Diversity of chromosomal karyotypes in maize and its relatives. Cytogenet Genome Res 129:6–16

Alfenito MR, Birchler JA (1993) Molecular characterization of a maize B chromosome centric sequence. Genetics 135:589–597

Ananiev EV, Phillips RL, Rines HW (1998) Chromosome-specific molecular organization of maize (*Zea mays L.*) centromeric regions. Proc Natl Acad Sci USA 95:13073–13078

Biggins S (2013) The composition, functions, and regulation of the budding yeast kinetochore. Genetics 194:817–846

Bilinski P, Distor K, Gutierrez-Lopez J, Mendoza GM, Shi J, Dawe RK, Ross-Ibarra J (2015) Diversity and evolution of centromere repeats in the maize genome. Chromosoma 124:57–65

Brown MT (1995) Sequence similarities between the yeast chromosome segregation protein Mif2 and the mammalian centromere protein CENP-C. Gene 160:111–116

Carlson W (1973) A procedure for localizing genetic factors controlling mitotic nondisjunction in the B chromosome of maize. Chromosoma 42:127–136

Carlson WR, Chou TS (1981) B chromosome nondisjunction in Corn: control by factors near the centromere. Genetics 97:379–389

Carroll CW, Milks KJ, Straight AF (2010) Dual recognition of CENP-A nucleosomes is required for centromere assembly. J Cell Biol 189:1143–1155

Cheeseman IM, Desai A (2008) Molecular architecture of the kinetochore–microtubule interface. Nat Rev Mol Cell Biol 9:33–46

Cheng ZK, Dong FG, Langdon T, Ouyang S, Buell CB, Gu MH, Blattner FR, Jiang JM (2002) Functional rice centromeres are marked by a satellite repeat and a

centromere-specific retrotransposon. Plant Cell 14:1691–1704

Da Ines O, Abe K, Goubely C, Gallego ME, White CI (2012) Differing requirements for RAD51 and DMC1 in meiotic pairing of centromeres and chromosome arms in *Arabidopsis thaliana*. PLoS Genet 8: e1002636

Dawe RK, Hiatt EN (2004) Plant neocentromeres: fast, focused, and driven. Chromosom Res 12:655–669

Dawe RK, Lowry EG, Gent JI, Stitzer MC, Higgins DM et al (2018) A kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. Cell 173:1–12

Dawe RK, Reed LM, Yu HG, Muszynski MG, Hiatt EN (1999) A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. Plant Cell 11:1227–1238

DeLuca JG, Gall WE, Ciferri C, Cimini D, Musacchio A, Salmon ED (2006) Kinetochore microtubule dynamics and attachment stability are regulated by Hec1. Cell 127:969–982

DeLuca JG, Musacchio A (2012) Structural organization of the kinetochore–microtubule interface. Curr Opin Cell Biol 24:48–56

Dong Q, Han F (2012) Phosphorylation of histone H2A is associated with centromere function and maintenance in meiosis. Plant J 71:800–809

Du Y, Dawe RK (2007) Maize NDC80 is a constitutive feature of the central kinetochore. Chromosom Res 15:767–775

Du Y, Topp CN, Dawe RK (2010) DNA binding of centromere protein C (CENP-C) is stabilized by single-stranded RNA. PLoS Genet 6:e1000835

Elowe S (2011) Bub1 and BubR1: At the interface between chromosome attachment and the spindle checkpoint. Mol Cell Biol 31:3085–3093

Fu S, Lv Z, Gao Z, Wu H, Pang J, Zhang B, Dong Q, Guo X, Wang X, Birchler JA (2013) De novo centromere formation on a chromosome fragment in maize. Proc Natl Acad Sci USA 110:6033–6036

Gao Z, Fu S, Dong Q, Han F, Birchler JA (2011) Inactivation of a centromere during the formation of a translocation in maize. Chromosom Res 19:755–761

Gent JI, Dawe RK (2012) RNA as a structural and regulatory component of the centromere. Annu Rev Genet 46:443–453

Gent JI, Dong Y, Jiang J, Dawe RK (2012) Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. Nucl Acids Res 40:1550–1560

Gent JI, Madzima TF, Bader R, Kent MR, Zhang X et al (2014) Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-Directed DNA methylation. Plant Cell 26:4903–4917

Gent JI, Schneider KL, Topp CN, Rodriguez C, Presting GG, Dawe RK (2011) Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning. Epigenetics Chromatin 4:3

Gent JI, Wang N, Dawe RK (2017) Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. Genome Biol 18:121

Gent JI, Wang K, Jiang J, Dawe RK (2015) Stable patterns of CENH3 occupancy through maize lineages containing genetically similar centromeres. Genetics 200:1105–1116

Gong ZY, Yu HX, Huang J, Yi CD, Gu MH (2009) Unstable transmission of rice chromosomes without functional centromeric repeats in asexual propagation. Chromosom Res 17:863–872

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES et al (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Han F, Gao Z, Birchler JA (2009) Reactivation of an inactive centromere reveals epigenetic and structural components for centromere specification in maize. Plant Cell 21:1929–1939

Han F, Gao Z, Yu W, Birchler JA (2007a) Minichromosome analysis of chromosome pairing, disjunction, and sister chromatid cohesion in maize. Plant Cell 19:3853–3863

Han F, Lamb JC, Birchler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. Proc Natl Acad Sci USA 103:3238–3243

Han F, Lamb JC, Yu W, Gao Z, Birchler JA (2007b) Centromere function and nondisjunction are independent components of the maize B chromosome accumulation mechanism. Plant Cell 19:524–533

Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N et al (2013) The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat Struct Mol Biol 20:687–695

Henikoff S, Ahmad K, Platero JS, van Steensel B (2000) Heterochromatic deposition of centromeric histone H3-like proteins. Proc Natl Acad Sci USA 97:716–721

Hiatt EN, Kentner EK, Dawe RK (2002) Independently regulated neocentromere activity of two classes of tandem repeat arrays. Plant Cell 14 (2):407–420

Hori T, Haraguchi T, Hiraoka Y, Kimura H, Fukagawa T (2003) Dynamic behavior of Nuf2-Hec1 complex that localizes to the centrosome and centromere and is essential for mitotic progression in vertebrate cells. J Cell Sci 116:3347–3362

Houben A, Wako T, Furushima-Shimogawara R, Presting G, Kunzel G, Schubert II, Fukui K (1999) Short communication: the cell cycle dependent phosphorylation of histone H3 is correlated with the condensation of plant mitotic chromosomes. Plant J 18:675–679

Jin W, Lamb JC, Vega JM, Dawe RK, Birchler JA, Jiang J (2005) Molecular and functional dissection of the maize B chromosome centromere. Plant Cell 17:1412–1423

Jin W, Lamb JC, Zhang W, Kolano B, Birchler JA, Jiang J (2008) Histone modifications associated with

both A and B chromosomes of maize. Chromosom Res 16:1203–1214
Jin W, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize centromeres: Organization and functional adaptation in the genetic background of oat. Plant Cell 16:571–581
Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC et al (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527
Kanizay LB, Albert PS, Birchler JA, Dawe RK (2013) Intragenomic conflict between the two major knob repeats of maize. Genetics 194:81–89
Kaszás E, Birchler JA (1996) Misdivision analysis of centromere structure in maize. EMBO J 15:5246–55
Kaszás E, Birchler JA (1998) Meiotic transmission rates correlate with physical features of rearranged centromeres in maize. Genetics 150:1683–1692
Kaszas E, Cande WZ (2000) Phosphorylation of histone H3 is correlated with changes in the maintenance of sister chromatid cohesion during meiosis in maize, rather than the condensation of the chromatin. J Cell Sci 113:3217–3226
Kato H, Jiang J, Zhou B, Rozendaal M, Feng H, Ghirlando R, Xiao TS, Straight AF, Bai Y (2013) A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. Science 340:1110–1113
Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci USA 101:13554–13559
Kato A, Zheng Y-Z, Auger DL, Phelps-Durr T, Bauer MJ, Lamb JC, Birchler JA (2005) Minichromosomes derived from the B chromosome of maize. Cytogenet Genome Res 109:156–165
Kelliher T, Starr D, Wang W, McCuiston J, Zhong H, Nuccio ML, Martin B (2016) Maternal haploids are preferentially induced by CENH3-tailswap transgenic complementation in maize. Front. Plant Sci 7:414
Kemp B, Boumil RM, Stewart MN, Dawson DS (2004) A role for centromere pairing in meiotic chromosome segregation. Genes Dev 18:1946–1951
Kline SL, Cheeseman IM, Hori T, Fukagawa T, Desai A (2006) The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation. J Cell Biol 173:9–17
Koo DH, Han FP, Birchler JA, Jiang JM (2011) Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. Genome Res 21:908–914
Koo DH, Zhao H, Jiang J (2016) Chromatin-associated transcripts of tandemly repetitive DNA sequences revealed by RNA-FISH. Chromosom Res 24:467–480
Kouzarides T (2007) Chromatin modifications and their function. Cell 128(4):693–705
Kynast RG, Riera-Lizarazu O, Vales MI, Okagaki RJ, Maquieira SB, Chen G, Ananiev EV, Odland WE, Russell CD, Stec AO, Livingston SM, Zaia HA, Rines HW, Phillips RL (2001) A complete set of maize individual chromosome additions to the oat genome. Plant Phys 125:1216–1227
Lamb JC, Birchler JA (2006) Retroelement genome painting: cytological visualization of retroelement expansions in the genera Zea and Tripsacum. Genetics 173:1007–1021
Lamb JC, Kato A, Birchler JA (2005) Sequences associated with a chromosome centromeres are present throughout the maize B chromosome. Chromosoma 113:337–349
Lamb JC, Meyer JM, Birchler JA (2007) A hemicentric inversion in the maize line knobless Tama flint created two sites of centromeric elements and moved the kinetochore-forming region. Chromosoma 116:237–247
Lara-Gonzalez P, Westhorpe FG, Taylor SS (2012) The spindle assembly checkpoint. Current Biol 22:R980
Lee HR, Zhang WL, Langdon T, Jin WW, Yan HH, Cheng ZK, Jiang JM (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc Natl Acad Sci USA 102:11793–11798
Li X, Dawe RK (2009) Fused sister kinetochores initiate the reductional division in meiosis I. Nat Cell Biol 11:1103–1108
Liu Y, Su H, Pang J, Gao Z, Wang XJ et al (2015) Sequential de novo centromere formation and inactivation on a chromosomal fragment in maize. Proc Natl Acad Sci USA 112:E1263–E1271
Liu Y, Su H, Liu Y, Zhang J, Dong Q, Birchler JA, Han F (2017) Cohesion and centromere activity are required for phosphorylation of histone H3 in maize. Plant J 92:1121–1131
Logarinho E, Resende T, Torres C, Bousbaa H (2008) The human spindle assembly checkpoint protein Bub3 is required for the establishment of efficient kinetochore–microtubule attachments. Mol Biol Cell 19:1798–1813
Maheshwari S, Ishii T, Brown CT, Houben A, Comai L (2017) Centromere location in Arabidopsis is unaltered by extreme divergence in CENH3 protein sequence. Genome Res 27:471–478
Maheshwari S, Tan EH, West A, Franklin FC, Comai L, Chan SW (2015) Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet 11:e1004970
Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. Nat Struct Mol Biol 10:882–891
Martinez-Perez E, Shaw P, Reader S, Aragon AL, Miller T, Moore G (1999) Homologous chromosome pairing in wheat. J Cell Sci 112:1761–1769
Masonbrink RE, Birchler JA (2012) Accumulation of multiple copies of maize minichromosomes. Cytogenet Genome Res 137:50–59
McClintock B (1939) The behavior of successive nuclear divisions of a chromosome broken at meiosis. Proc Natl Acad Sci USA 25:405–416

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, Garcia JF, DeRisi JL, Smith T, Tobias C, Ross-Ibarra J, Korf I, Chan SW (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14: R10

Meluh PB, Koshland D (1995) Evidence that the MIF2 gene of *Saccharomyces cerevisiae* encodes a centromere protein with homology to the mammalian centromere protein CENP-C. Mol Biol Cell 6:793–807

Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. Genome Biol 7:R23

Mondin M, Santos-Serejo JA, Bertão MR, Laborda P, Pizzaia D et al (2014) Karyotype variability in tropical maize sister inbred lines and hybrids compared with KYS standard line. Front Plant Sci 5:544

Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics 150:1615–1623

Musacchio A, Salmon ED (2007) The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol 8:379–393

Nasuda S, Hudakova S, Schubert I, Houben A, Endo TR (2005) Stable barley chromosomes without centromeric repeats. Proc Natl Acad Sci USA 102:9842–9847

Neumann P, Navrátilová A, Koblížková A, Kejnovský E, Hřibová E et al (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA 2:4

Osman F, Whitby MC (2013) Emerging roles for centromere-associated proteins in DNA repair and genetic recombination. Biochem Soc Trans 41:1726–1730

Page BT, Wanous MK, Birchler JA (2001) Characterization of a maize chromosome 4 centromeric sequence: evidence for an evolutionary relationship with the B chromosome centromere. Genetics 159:291–302

Paterson AH, Bower JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Phillips D, Nibau C, Wnetrzak J, Jenkins G (2012) High resolution analysis of meiotic chromosome structure and behaviour in barley (*Hordeum vulgare* L.). PLoS ONE 7:e39539

Presting GG, Malysheva L, Fuchs J, Schubert I (1998) TY3/GYPSYretrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16 (6):721–728

Przewloka MR, Glover DM (2009) The kinetochore and the centromere: a working long distance relationship. Annu Rev Genet 43:439–465

Rines HW, Phillips RL, Kynast RG, Okagaki RJ, Galatowitsch MW, Huettl PA, Stec AO, Jacobs MS, Suresh J, Porter HL, Walch MD, Cabral CB (2009) Addition of individual chromosomes of maize inbreds B73 and Mo17 to oat cultivars Starter and Sun II: maized chromosomes retention, transmission, and plant phenotype. Theor Appl Geneti 119:1255–1264

Ravi M, Chan SW (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464:615–618

Rhoades MM (1942) Preferential segregation in maize. Genetics 27:395–407

Rhoades MM (1952) In: JW Gowen (ed) Preferential segregation in maize, in *Heterosis*. Iowa State College Press

Rhoades MM, Vilkomerson H (1942) On the anaphase movement of chromosomes. Proc Natl Acad Sci USA 28:433–436

Roman H (1947) Mitotic nondisjunction in the case of interchanges involving the B-type chromosome in maize. Genetics 32:391409

Roman H (1948) Directed fertilization in maize. Proc Natl Acad Sci USA 34:36–42

Ronceret A, Doutriaux MP, Golubovskaya IN, Pawlowski WP (2009) PHS1 regulates meiotic recombination and homologous chromosome pairing by controlling the transport of RAD50 to the nucleus. Proc Natl Acad Sci USA 106:20121–20126

Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves TA et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schneider KL, Xie ZD, Wolfgruber TK, Presting GG (2016) Inbreeding drives maize centromere evolution. Proc Natl Acad Sci USA 113:E987–E996

Sharma A, Presting GG (2014) Evolution of centromeric retrotransposons in grasses. Genome Biol Evol 6:1335–1352

Shi J, Wolf SE, Burke JM, Presting GG, Ross-Ibarra J, Dawe RK (2010) Widespread gene conversion in centromere cores. PLoS Biol 8:e1000327

Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. Trends Biochem Sci 24:181–185

Su H, Liu Y, Dong Q, Feng C, Zhang J, Liu Y, Birchler JA, Han F (2017) Dynamic location changes of Bub1-phosphorylated-H2AThr133 with CENH3 nucleosome in maize centromeric regions. New Phytol 214:682–694

Su H, Liu Y, Liu YX, Lv Z, Li H, Xie S et al (2016) Dynamic chromatin changes associated with de novo centromere formation in maize euchromatin. Plant J 88:854–866

Takeo S, Lake CM, Morais-de-Sá E, Sunkel CE, Hawley RS (2011) Synaptonemal complex-dependent

centromeric clustering and the initiation of synapsis in *Drosophila* oocytes. Curr Biol 21:1845–1851

Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an arabidopsis histone H3 variant. Plant Cell 14:1053–1066

Topp CN, Okagaki RJ, Melo JR, Kynast RG, Phillips RL, Dawe RK (2009) Identification of a maize neocentromere in an oat-maize addition line. Cytogenet Genome Res 124:228–238

Topp CN, Zhong CX, Dawe RK (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci USA 101:15986–15991

Trifonov EN, Nibhani R (2015) Review fifteen years of search for strong nucleosomes. Biopolymers 103:432–437

Tsubouchi T, Roeder GS (2005) A synaptonemal complex protein promotes homology-independent centromere coupling. Science 308:870–873

Unhavaithaya Y, Orr-Weaver TL (2013) Centromere proteins CENP-C and CAL1 functionally interact in meiosis for centromere clustering, pairing, and chromsome segregation. Proc Natl Acad Sci USA 110:19878–19883

Varma D, Salmon ED (2012) The KMN protein network–chief conductors of the kinetochore orchestra. J Cell Sci 125:5927–5936

Wang H, Bennetzen JL (2012) Centromere retention and loss during the descent of maize from a tetraploid ancestor. Proc Natl Acad Sci USA 109:21004–21009

Wang N, Dawe RK (2018) Centromere size and its relationship to haploid formation in plants. Mol Plant 11:398–406

Wang K, Wu Y, Zhang W, Dawe RK, Jiang J (2014) Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Res 24:107–116

Ward E (1973) Nondisjunction: localization of the controlling site in the maize B chromosome. Genetics 73:387–391

Wear EE, Song J, Zynda GJ, LeBlanc C, Lee TJ, Mickelson-Young L, Concia L, Mulvaney P, Szymanski ES, Allen GC, Martienssen RA, Vaugh MW, Hanley-Bowdoin L, Thompson WF (2017) Genomic analysis of the DNA replication timing program during mitotic S phase in maize (*Zea mays*) root tips. Plant Cell 29:2126–2149

Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S, Fuks G, Sanchez-Villeda H, Schroeder S, Fang Z, McMullen M, Davis G, Bowers JE, Paterson AH, Schaeffer M, Gardiner J, Cone K, Messing J, Soderlund C, Wing RA (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet 3:1254–1263

Wen R, Moore G, Shaw PJ (2012) Centromeres cluster de novo at the beginning of meiosis in *Brachypodium distachyon*. PLoS ONE 7:e44681

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130

Wolfgruber TK, Nakashima MM, Schneider KL, Sharma A, Xie Z, Albert PS, Xu R, Bilinski P, Dawe RK, Ross-Ibarra J, Birchler JA, Presting GG (2016) High quality maize centromere 10 sequence reveals evidence of frequent recombination events. Front Plant Sci. 7:308

Wolfgruber TK, Sharma A, Schneider KL, Albert PS, Koo DH, Shi J, Gao Z, Han F, Lee H, Xu R, Allison J, Birchler JA, Jiang J, Dawe RK, Presting GG (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. PLoS Genet 5: e1000743

Xiong W, Dooner HK, Du C (2016) Rolling-circle amplification of centromeric Helitrons in plant genomes. Plant J 88:1038–1045

Yang X, Zhao H, Zhang T, Zeng Z, Zhang P, Zhu B, Han Y, Braz GT, Casler MD, Schmutz J, Jiang J (2018) Amplification and adaptation of centromeric repeats in polyploid switchgrass species. New Phytol 218 (4):1645–1657

Yu HG, Muszynski MG, Kelly Dawe R (1999) The maize homologue of the cell cycle checkpoint protein MAD2 reveals kinetochore substructure and contrasting mitotic and meiotic localization patterns. J Cell Biol 145:425–435

Zhang W, Lee HR, Koo DH, Jiang J (2008) Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. Plant Cell 20:25–34

Zhang X, Li X, Marshall JB, Zhong CX, Dawe RK (2005) Phosphoserines on maize CENTROMERIC HISTONE H3 and histone H3 demarcate the centromere and pericentromere during chromosome segregation. Plant Cell 17:572–583

Zhang B, Lv Z, Pang J, Liu Y, Guo X, Fu S, Li J, Dong Q, Wu HJ, Gao Z, Wang XJ, Han F (2013a) Formation of a functional maize centromere after loss of centromeric sequences and gain of ectopic sequences. Plant Cell 25:1979–1989

Zhang J, Pawlowski WP, Han F (2013b) Centromere pairing in early meiotic prophase requires active centromeres and precedes installation of the synaptonemal complex in maize. Plant Cell 25:3900–3909

Zhang T, Talbert PB, Zhang W, Wu Y, Yang Z et al (2013c) The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. Proc Natl Acad Sci USA 110:E4875–E4883

Zhao X, Xu X, Xie H, Chen S, Jin W (2013) Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers. Plant Physiol 163:721–731

Zhao HN, Zeng ZX, Koo D-H, Gill BS, Birchler JA, Jiang JM (2017) Recurrent establishment of de novo centromeres in the pericentromeric region of maize chromosome 3. Chromosom Res 25:299–311

Zhao H, Zhu X, Wang K, Gent JI, Zhang W et al (2016) Gene expression and chromatin modifications associated with maize centromeres. G3: Genes Genomes Genet. 6:183–192

Zheng YZ, Roseman RR, Carlson WR (1999) Time course study of the chromosome-type breakage-fusion-bridge cycle in maize. Genetics 153:1435–1444

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

The Maize Methylome

6

Jaclyn M. Noshay, Peter A. Crisp and Nathan M. Springer

Abstract

DNA methylation is a chromatin modification that has generally been associated with gene silencing or heterochromatin. Plants have mechanisms to allow for the stable inheritance of DNA methylation through mitosis or meiosis. This creates the potential for DNA methylation to provide epigenetic inheritance for traits in maize and other crops. Epigenetics refers to heritable transmission of information that is not solely attributable to DNA sequence. Several examples of epigenetic inheritance were first described in maize including paramutation, imprinting, and transposable element inactivation. There is evidence that DNA methylation is associated with each of these epigenetic phenomena. In addition, natural variation for epigenetic states may contribute substantially to variation among maize inbreds and could be an important source of variation for crop improvement. Advances in our understanding of the molecular mechanisms controlling DNA methylation in Arabidopsis have provided clues to the genes and pathways likely to be important in maize. Recent technological developments have provided the opportunity to characterize the genome-wide distribution of DNA methylation in the maize genome. This has provided insights into the patterns of DNA methylation in plant species with large, complex genomes and has led to the identification of potential cryptic genomic information that is silenced by DNA methylation. We will summarize current understanding of the mechanisms that regulate methylation and factors that influence variation and stability of the maize methylome.

6.1 Introduction

In maize, as in other eukaryotes, DNA methylation refers to the addition of a methyl group to the 5′ carbon of cytosine residues. This methyl group is added after DNA replication. Therefore, the faithful maintenance of DNA methylation patterns requires mechanisms to copy DNA methylation onto the daughter strand. A large majority of DNA methylation in maize, and other plants, is found at CG or CHG (where H is any base except G) sites that have symmetry across the two strands of DNA (Niederhuth et al. 2016). This allows for the maintenance of DNA methylation through targeted methylation of hemi-methylated DNA that results from the incorporation of unmethylated cytosines during DNA replication. Cytosine residues that are not

J. M. Noshay · P. A. Crisp · N. M. Springer (✉)
Department of Plant and Microbial Biology,
University of Minnesota, Saint Paul,
MN 55108, USA
e-mail: springer@umn.edu

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_6

followed by a G in the next two bases (CHH sites) can also be methylated but require alternative mechanisms for maintenance of the patterns following replication (Law and Jacobsen 2010; Matzke and Mosher 2014; Springer and Schmitz 2017). In recent years, Arabidopsis has provided a model system for studying DNA methylation due to the availability of reverse genetics resources and the viability of mutants with severely reduced DNA methylation (Law and Jacobsen 2010; Matzke and Mosher 2014). Our knowledge of the specific mechanisms that control DNA methylation and the role of DNA methylation in maize and other crop plants is more limited. Here, we will describe what is known in maize and contrast with data from Arabidopsis noting both conserved features and key differences.

6.2 Methods for Documenting DNA Methylation

There are a variety of approaches that have been utilized to monitor DNA methylation, with varying levels of sensitivity and specificity (reviewed by Zilberman et al. 2007). The genomic proportion of cytosine residues that are methylated can be roughly estimated by HPLC (Papa 2001). This approach is useful for quantifying genome-wide DNA methylation levels, but it cannot determine the level of methylation at specific sequence contexts, sites, or regions in the genome. In many cases, the presence of DNA methylation can inhibit digestion by restriction enzymes, and in some rare cases, there are restriction enzymes (McrBC, FspEI, MspJI) that require DNA methylation in order to cut a site (Loenen and Raleigh 2014). These methylation-sensitive or methylation-dependent enzymes can be combined with Southern blotting or quantitative PCR approaches to document the presence or absence of methylation at specific sites in the genome (Zhang et al. 2014). In general, the use of restriction enzymes for surveying DNA methylation can provide data for specific sites but tends to be only partially quantitative and can be difficult to apply in a high-throughput fashion. Methylation-sensitive enzymes can be combined with AFLP-based approaches to provide a survey of methylation at many different sites (Lu et al. 2008). Methylation-dependent enzymes have been used in combination with shotgun sequencing or microarray approaches for genome-wide identification of unmethylated regions referred to as methylation filtration (Palmer 2003; Rabinowicz et al. 2005). Another approach for documenting genome-wide methylation levels utilizes a 5-methylcytosine antibody for immunoprecipitation of methylated DNA (meDIP) (Eichten et al. 2011). This approach enriches for fragments containing DNA methylation and can be combined with microarrays or high-throughput sequencing approaches to provide genome-wide profiles. The methylation filtration and meDIP assess regional methylation throughout a genome but do not provide single-base resolution of DNA methylation.

The "gold-standard" approach for measuring DNA methylation is with sodium bisulfite treatment followed by sequencing (Lister et al. 2008). Treatment of single-stranded DNA with sodium bisulfite will result in conversion of unmethylated cytosine residues to uracil, while methylated cytosines are not converted. Sequencing of treated molecules reveals which bases remained as cytosine (methylated) and which bases were converted (unmethylated). By sequencing multiple molecules, the frequency of methylation at any particular site can be determined. This approach was initially combined with PCR to document methylation at particular genomic regions. In recent years, this has been paired with next-generation sequencing to perform whole-genome bisulfite sequencing (WGBS) which provides base-level resolution and context-specific information for DNA methylation throughout the portion of the genome for which unique alignments are possible (Lister et al. 2008; Regulski et al. 2013; Gent et al. 2013). Bisulfite treatment can also be paired with sequence capture approaches to provide single-base resolution for a subset of genomic regions (Li et al. 2015c).

6.3 Genomic Distribution of DNA Methylation in Maize

WGBS has been used to document the genome-wide distribution of DNA methylation in maize (Regulski et al. 2013; Gent et al. 2013; West et al. 2014). However, it is worth noting that current short-read sequencing and bioinformatics approaches cannot interrogate the entire genome. WGBS allows analysis of regions covered by uniquely aligning reads, which results in coverage for ~70% of the maize genome. Genic (78% coverage) and intergenic (90% coverage) regions have substantially higher coverage than TEs (60% coverage) for methylation data (Fig. 6.1). WGBS profiles have revealed that plant genomes have similar mechanisms for DNA methylation, but the frequency and patterning of methylation domains varies among species (Niederhuth et al. 2016; Springer and Schmitz 2017). While maize has most of the methylation machinery found in Arabidopsis, it must operate to methylate a genome with a different organization. Arabidopsis has a relatively small genome with a high gene density (The Arabidopsis Genome Initiative 2000), most genes are not located near TEs, and the vast majority of heterochromatin in the Arabidopsis genome is located in pericentromeric regions. In contrast, the maize genome has a much lower gene density (Schnable et al. 2009; Jiao et al. 2017) and TEs are prevalent throughout the whole length of maize chromosomes (Baucom et al. 2009). The total abundance and relative distribution of CG, CHG, and CHH across the genomes of Arabidopsis and maize are distinct (West et al. 2014; Neiderhuth et al. 2016). Methylation in all three sequence contexts is highly enriched within TEs, repeat sequences, and pericentromeric regions in Arabidopsis (West et al. 2014). Maize has among the highest levels of CG and CHG methylation in species with characterized methylation profiles, and methylation at CG and CHG contexts are found throughout the length of the maize chromosome (Springer and Schmitz 2017; West et al. 2014; Nierderhuth et al. 2016). In contrast, the levels of CHH methylation in maize are relatively low compared to many other plant species (West et al. 2014; Niederhuth et al. 2016).

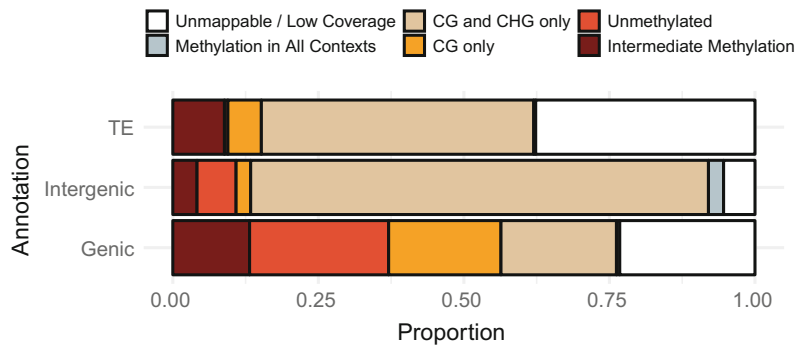

Fig. 6.1 Frequency of methylation domains in different genomic regions. A WGBS dataset for maize earshoot (Li et al. 2015a, b, c) was mapped to version 4 of the maize B73 genome (Jiao et al. 2017). The level of DNA methylation in each sequence context was determined for each 100 bp region as described in West et al. (2014). Each 100 bp region was classified as genic (7.3% of genome), TE (72.3%), or intergenic (20.4%) based on B73v4 annotations. Each 100 bp region was classified into one of six groups using the following criteria: unmappable/low coverage (regions with <2X coverage), all contexts methylated (>15% CHH methylation), CG/CHG only (40% CG and >40% CHG), CG only (>40% CG but <40% CHG), unmethylated (<10% methylation in all sequence contexts), and intermediate methylation (sufficient coverage but not classified as one of the other groups). The proportion of 100 bp regions for each subset of annotated features were determined

6.3.1 Methylation Domains in the Maize Genome

Assessing the relative levels of CG, CHG, and CHH methylation in windows of the maize genome can be used to define different types of methylation domains (Springer and Schmitz 2017). The methylation domains include CG/CHG/CHH regions (RNA-directed DNA methylation, or RdDM targets), CG/CHG regions (heterochromatin), CG only (gene body methylation—gBM), unmethylated regions, and unclassified regions with intermediate levels of DNA methylation (Fig. 6.1). CG/CHG domains, which contain high levels of CG and CHG methylation, but very low levels of CHH methylation, are the most common type in the maize genome, accounting for large portions of intergenic and TE regions of the genome but are less abundant within genes (Fig. 6.1). The RdDM targets, which have elevated methylation in all three contexts, only account for 2% of the maize genome and are most prevalent within intergenic regions. Regions with only CG methylation account for ~6% of the maize genome and are often found within maize gene bodies. Approximately 11% of the maize genome has low levels of methylation in all three contexts, and this is most prevalent within the genic portions of the maize genome and is quite rare in TEs. Another 10% of the maize genome has intermediate levels of DNA methylation that are difficult to classify.

6.3.2 DNA Methylation Patterns at Maize Genes

The distribution of methylation within plant genomes reflects the distinct methylation profiles at genes and TEs. In general, CG and CHG methylation levels are high in non-genic regions but drop to low levels near the transcription start site (TSS) and transcription termination site (TTS) of annotated genes (Regulski et al. 2013; Gent et al. 2013; West et al. 2014). Within gene bodies, there are moderate levels of CG methylation likely reflecting gene body methylation (Neiderhuth et al. 2016). Maize also contains significant levels of CHG methylation in gene bodies that is partially attributable to methylation of TEs found within introns (West et al. 2014). CHH methylation is enriched in regions flanking maize genes (Gent et al. 2013). These mCHH islands mark the boundary between high levels of CG and CHG methylation outside of maize genes and the reduced levels of methylation in genes (Li et al. 2015a). Several factors influence the profile of DNA methylation over maize genes. In general, highly expressed genes have the lowest levels of DNA methylation at the TSS and TTS (Regulski et al. 2013; Gent et al. 2013; West et al. 2014). However, the inverse pattern is observed for CHH methylation in regions upstream of the TSS (Gent et al. 2013). Genes located in syntenic positions relative to other grasses exhibit much lower levels of DNA methylation than inserted (non-syntenic) genes (Eichten et al. 2011; West et al. 2014). There is no evidence for differential levels of DNA methylation for genes in the two subgenomes that have resulted from the ancient whole-genome duplication event in maize (Eichten et al. 2011; West et al. 2014).

6.3.3 DNA Methylation Patterns at Maize TEs

DNA methylation at TEs is high relative to flanking regions (West et al. 2014). The levels of CG and CHG methylation over TEs are higher in maize than in Arabidopsis (West et al. 2014), with more gradual transitions from low to high methylation levels at the edges of TEs, suggesting greater spreading of DNA methylation from TEs to flanking regions in maize (Eichten et al. 2012). The analysis of transposon superfamilies revealed variation in chromatin profiles (West et al. 2014). While CG and CHG methylation are very high for all families, there is variation for the level of CHH methylation and H3K9me2 (West et al. 2014). There is also evidence for family-specific variation in whether DNA methylation can spread to flanking regions, suggesting that TE families associated with

spreading are more likely to reduce the expression of nearby genes than families without spreading (Eichten et al. 2012). The association of CG and CHG methylation (inactive transcription) with spreading retrotransposon families and CHH (active transcription) with non-spreading retrotransposon families can explain this gene expression correlation. The methylation levels of transposons located within maize genes are quite similar to the levels for intergenic TEs even though these regions undergo active transcription (West et al. 2014). This suggests that methylated TEs do not pose a barrier to transcriptional elongation. However, there is evidence that plants require machinery to allow for proper transcription and splicing of regions that are highly methylated (To et al. 2015).

6.4 Molecular Mechanisms Regulating DNA Methylation

DNA methylation at any locus is influenced by a variety of processes including methylation maintenance, de novo methylation, and demethylation. We will describe the mechanisms expected to control CG, CHG, and CHH methylation based on studies in Arabidopsis and the evidence for similar systems being present in maize. The Arabidopsis genome encodes seven DNA methyltransferases including *DOMAINS REARRANGED METHYLASE 2* (*DRM1*) and *DRM2*, *CHROMOMETHYLASE 1* (*CMT1*), *CMT2*, and *CMT3*, *METHYLTRANSFERASE 1* (*MET1*) and *MET2*. Four of these methyltransferases (*DRM2*, *CMT2*, *CMT3*, and *MET1*) are responsible for the bulk of methylation in Arabidopsis and contribute to different maintenance and de novo methylation pathways (Stroud et al. 2013; Law and Jacobsen 2010; Matzke and Mosher 2014; Du et al. 2015).

6.4.1 CG Methylation

Genetic analysis has shown that *MET1* is required for CG methylation maintenance in Arabidopsis (Law and Jacobsen 2010). MET1 is dependent on three VARIANTS IN METHYLATION (VIM) proteins, which are ubiquitin E3 ligases containing an SRA domain that binds hemi-methylated DNA (Du et al. 2015; Woo et al. 2008). After MET1 is recruited to hemi-methylated CG sites, it functions to methylate the opposing strand, providing a robust mechanism to transmit CG methylation patterns following DNA replication. In maize, two tandem duplicates of MET1-like genes (*Zmet1*—Zm00001d018976 and Zm00001d018977) have been identified (Li et al. 2014a). The maize genome also encodes at least three VIM1-like genes. The tandemly duplicated MET1-like genes in maize likely play critical roles in maintaining CG methylation in the maize genome similar to *MET1* in Arabidopsis. To date, loss-of-function alleles has not been isolated for these genes through forward or reverse genetics approaches, limiting functional studies of these genes in maize.

6.4.2 CHG Methylation

In Arabidopsis, the bulk of CHG methylation is maintained by the chromomethylase *CMT3* (Matzke and Mosher 2014; Du et al. 2015; Bewick et al. 2016). CMT3 contains a BAH domain, a DNA methyltransferase domain, and a chromodomain. The chromodomain and BAH domain provide the ability for CMT3 to bind to histone H3 that has dimethylated lysine 9 (H3K9me2) (Du et al. 2012). In Arabidopsis, the enzyme that provides H3K9me2, KRYPTONITE (KYP), binds to CHG methylation (Du et al. 2014). This provides a self-reinforcing loop between CHG DNA methylation and H3K9me2 which provides a mechanism for stable memory of this chromatin state (Du et al. 2015). The maize genome encodes two paralogs that are related to Arabidopsis *CMT3*; *Zmet2* (*Dmt102*—Zm00001d026291); and *Zmet5* (*Dmt105*—Zm00001d002330) (Papa 2001; Makarevitch et al. 2007). A loss-of-function allele, *zmet2-m1*, results in significant reductions of genomic CHG methylation levels (Papa 2001). Other partial

loss-of-function alleles for *zmet2* or *zmet5* also result in partial reductions in CHG methylation in maize (Li et al. 2014a). *Zmet2* and *Zmet5* are expressed in similar patterns across a variety of tissues in B73 with slightly higher expression seen in *Zmet2* (Li et al. 2014a). Attempts to isolate plants homozygous for mutations in both *Zmet2* and *Zmet5* were unsuccessful, suggesting essential functions for CHG methylation in maize (Li et al. 2014a). Recent work suggests that the vast majority of "CHG" methylation in plant genomes is confined to CWG (where W is A or T) sites with very little methylation of the external C of CCG sites (Gouil and Baulcombe 2016).

6.4.3 CHH Methylation

There is evidence for two separate pathways for maintaining CHH methylation in plant genomes. The RdDM, involving *DRM1* and *DRM2*, plays an important role in methylation of CHH, particularly in genomic regions near genes (Law and Jacobsen 2010; Matzke and Mosher 2014). RdDM involves the production and perception of 24nt siRNAs through the combined action of two plant-specific RNA polymerases, PolIV and PolV as well as RNA-dependent RNA polymerase RDR2 and additional components (Matzke and Mosher 2014). The recruitment of RdDM activity to specific loci appears to require the presence of DNA methylation and specific chromatin modifications, suggesting that RdDM plays a critical role in maintaining CHH methylation patterns but may not actually represent true de novo methylation activities (Law et al. 2013; Greenberg et al. 2013; Johnson et al. 2014). True de novo methylation activities may require the activity of 21nt siRNAs with AGO6 and RDR6 to recruit DRM2 to specific target loci (Panda and Slotkin 2013; McCue et al. 2014). Arabidopsis also encodes a third domain rearranged methyltransferase, *DRM3* (Henderson et al. 2010). Interestingly, although the DRM3 protein is catalytically inactive due to changes in the active site, it is a required cofactor for proper activity of DRM2 (Henderson et al. 2010).

In addition to DRM-dependent CHH methylation targeted by RdDM activities, there is also evidence for CHH methylation in deep heterochromatin that requires the chromomethylase *CMT2* (Zemach et al. 2013). These regions are likely inaccessible to PolIV/PolV activity and instead depend on CHH methylation activities from CMT2 (Stroud et al. 2014). In order to methylate these regions, CMT2 is recruited by histone methylation (Du et al. 2015). This "CHH" methylation appears to be largely confined to CWA (where W is A or T) sites (Gouil and Baulcombe 2016). Together, RdDM (utilizing DRM activities) and CMT2 maintain CHH methylation in the Arabidopsis genome.

Maize contains several DRM-like genes including *Zmet3* (*Dmt103*—Zm00001d048516), *Zmet6* (*Dmt106*—Zm00001d010928), and *Zmet7* (*Dmt107*—Zm00001d027329). *Zmet3* and *Zmet7* are retained duplicates most closely related to DRM1/2, and *Zmet6* is most similar to DRM3 (Li et al. 2014a). *Zmet3* and *Zmet7* are likely retained duplicates arising from a whole-genome duplication event in maize and exhibit similar expression patterns throughout development (Li et al. 2014a). Two loss-of-function alleles have been recovered for *Zmet7* (Li et al. 2014a), but there are no documented loss-of-function alleles for *Zmet3* to date. Mutations in *Zmet7* do not have significant effects on CHH methylation in maize, but this could be due to redundancy with *Zmet3* (Li et al. 2014a). The *Zmet6* gene encodes a protein predicted to be catalytically inactive, similar to DRM3 due to changes in the amino acid sequence near the active site of the methyltransferase domain. Maize also encodes orthologs for many of the components of the RdDM pathway (Haag et al. 2014). Several of these genes have been identified through forward genetics that identified genes required for paramutation at *R* or *Pl* (Alleman et al. 2006; Stonaker et al. 2009; Hollick 2017). Mutations in several of these genes have been shown to be required for maintaining CHH methylation at genomic regions with high (>20%) levels of CHH methylation (Li et al. 2014a, 2015a). These mutants that eliminate regions of high CHH methylation have relatively minor

effects on gene expression in maize (Forestan et al. 2017; Anderson et al. 2018). Interestingly, unlike other grasses, maize does not contain *CMT2* orthologs (Zemach et al. 2013; Bewick et al. 2016). In maize, the deep heterochromatin regions are marked with high levels of CG and CHG methylation but low (~1–5%) levels of CHH methylation (Li et al. 2014a) that is largely confined to CWA sites (Gouil and Baulcombe 2016). It appears that this CHH methylation may depend on CHH activities of *Zmet2*/*Zmet5* (Li et al. 2014a; Gouil and Baulcombe 2016).

6.4.4 Demethylation

While plant genomes have encoded proteins that contribute to a variety of pathways to catalyze DNA methylation, they also encode enzymes capable of active demethylation (Zhang and Zhu 2012). Demethylation is essential for certain plant developmental processes, for instance tomato fruit ripening (Liu et al. 2015) and imprinting (Bauer and Fischer 2011). Passive demethylation occurs via the failure to methylate hemi-methylated molecules that are present following DNA replication. Active demethylation (Zhang and Zhu 2012) occurs through targeted glycosylase activities. Arabidopsis includes at least four related genes including *DEMETER* (*DME*), *REPRESSOR OF SILENCING* 1 (*ROS1*), *DEMETER-LIKE 2* (*DML2*) and *DML3* that are DNA glycosylases responsible for removal of methylated cytosines through a base-excision-repair mechanism (Zhang and Zhu 2012). The maize genome encodes several DNA glycosylases (DNGs) that are homologous to those in Arabidopsis, including a DME-like gene (Zm00001d016516) and *ROS1* homologs *dng101* (Zm00001d053251) and *dng103* (Zm00001d038302), but no loss-of-function alleles for these genes have been reported. We still have a limited understanding of the mechanisms that target these demethylation activities to specific genomic regions, but there is clear evidence that the existing methylation patterns in the Arabidopsis genome reflect a balance of methylating and demethylating activities.

6.5 Sources of Variation for the Maize Methylome and Inheritance

Understanding the frequency and distribution of differentially methylated regions (DMRs) among maize genotypes could help connect DNA methylation with phenotypic variation. In addition, understanding whether changes occur stochastically, during development, or in response to the environment is important for documenting the stability of DNA methylation. We also must understand the inheritance of variation to determine whether DNA methylation has the potential to influence heritability of traits and how to account for DNA methylation in genomic selection models or GWAS.

6.5.1 Mechanisms of Variation

Multiple mechanisms have been proposed to give rise to variation in DNA methylation, from pure epialleles with no genetic changes to obligatory and facilitated epialleles that depend on underlying genetic variation (Richards 2006). Examples of pure epialleles (Eichten et al. 2011) and of epialleles associated with genetic changes (Eichten et al. 2012) have been reported in maize. Given that >60% of the maize genome is annotated as transposable elements (Schnable 2009; Jiao et al. 2017), and that the composition and organization of TEs can vary greatly between inbred lines (Wang et al. 2015), this genetic variation may underpin a significant portion of variation in the methylome.

The rate of spontaneous epimutations has been studied in detail in Arabidopsis using mutation accumulation lines. Such investigations have focused on DMRs rather than single methylation polymorphisms (SMPs) because regional changes in DNA methylation are likely more functionally relevant. DMRs arise at rates comparable to genetic mutations such as SNPs (Schmitz et al. 2011; Becker et al. 2011). However, the frequency of epimutations at single cytosine residues, SMPs, is many orders of magnitude more frequent (Becker et al. 2011). It

is likely different regions of the epigenome and different methylation contexts vary in SMP rates (van der Graaf et al. 2015). Transgenerational studies in Arabidopsis highlight two significant points; SMPs can occur stochastically and SMPs are reversible, in contrast to genetic mutation. Thus, some variation in the DNA methylome arises over time through random stochastic variation. Such variation does not increase linearly with time indicating that such changes, while often stable and heritable, are also reversible. However, there was less evidence for high rates of reversible changes in methylation on a regional level (DMRs) in these studies.

6.5.2 Sources of Variation

Multiple studies employing a variety of technologies have demonstrated natural variation for DNA methylation in maize (Makarevitch et al. 2007; Eichten et al. 2011, 2013; Regulski et al. 2013; Li et al. 2014b, 2015b). Initial efforts identified around 700 DMRs using meDIP between B73 and Mo17 (Eichten et al. 2011). A larger scan that included ~50 diverse maize inbreds identified 1,966 common and 1,754 rare DMRs (Eichten et al. 2013). A shift from meDIP to WGBS greatly increased the number of context-specific DMRs that were identified in maize, with 5,000–20,000 context-specific DMRs between any two genotypes (Li et al. 2015b).

When considering this extensive epigenomic variation, it is important to consider the background genetic variation. Many DMRs can be associated with local genomic variation (Eichten et al. 2011, 2013). For instance; Eichten et al. (2013) reported that half of the common DMRs assessed in a panel of 50 inbred lines were associated with SNPs found within or near the DMRs; Li et al. (2015b) found that the majority of DMRs were associated with local sequence variation. These studies highlight the strong relationship between genetic and epigenetic variations. Nevertheless, examples of DMRs occurring in genomic regions that are apparently identical between inbreds (e.g., B73 and Mo17) indicate the existence of pure epialleles (Eichten et al. 2011; Li et al. 2015b). Overall, most studies have found greater than 99% of the methylome is conserved within a species (Li et al. 2015b). Yet, this leaves ample variation at hundreds to thousands of loci, which may contribute to phenotypic variation and breeding outcomes.

Given that DNA methylation variation can potentially occur more rapidly than genomic variation and that it is reversible, regulation of the methylome may provide a means for local and rapid acclimation or adaptation to new environments. Despite this attractive hypothesis, few concrete documented examples of environmentally induced, heritable changes in DNA methylation exist (Pecinka and Scheid 2012; Crisp et al. 2016). Profiling of maize plants subjected to heat, cold, and UV revealed no evidence for consistent changes in DNA methylation in response to stress (Eichten and Springer 2015). This analysis also found that stress did not appear to increase the rate of epimutation. The examples of variation that have been identified tend to be enriched in the CHH context and lack stable inheritance patterns (Secco et al. 2015). The emerging trend that the methylome is largely impervious to environmental perturbation has important implications for breeding, allowing selection for epigenetic traits for large-scale agricultural application where plants can be grown under a wide variety of environments.

Another potential source of DNA methylation variation is developmental and cellular differentiation leading to cell-type- or tissue-specific variation. In animals, there are well-documented examples of developmental epigenetic variation (Feng et al. 2010; Heard and Martienssen 2014). Similarly, maize endosperm and embryo have a number of differences in DNA methylation (Wang et al. 2015), consistent with findings in rice and Arabidopsis (Gehring et al. 2009; Hsieh et al. 2009; Zemach et al. 2010). In the endosperm, there is widespread hypomethylation of the maternal genome, particularly at TEs, associated with the activation of endosperm-specific DNA demethylases (Wang et al. 2015). Another example of a cell-type-specific methylome regulation occurs in the columella. The columella in

the Arabidopsis root cap has been identified as the most hypomethylated Arabidopsis cell/tissue to date (Kawakatsu et al. 2016). Similarly, developmental regulation of DNA methylation appears to play an essential role in tomato fruit ripening, where specific gene promoters become hypomethylated during the progressive stages of ripening (Zong et al. 2013). Notwithstanding these notable examples of DNA methylation in certain tissues there is very little evidence for variation in DNA methylation between most cell types and during the majority of vegetative development (Kawakatsu et al. 2016).

In contrast to abiotic stress and development, it has long been known that tissue culture induces a remarkable degree of variation in the methylome (Kaeppler and Phillips 1993). The tissue culture process represents a traumatic genomic stress to plant cells (Phillips et al. 1994; Kaeppler et al. 2000). Despite the expectation that plants regenerated from tissue culture will be clones, regenerates often display heritable phenotypic and molecular variation (Stelpflug et al. 2014). Methylome profiles of regenerated plants have identified 479 DMRs compared to siblings not subjected to tissue culture, with a bias toward hypomethylation (Stelpflug et al. 2014). Of the hypomethylated loci, 75% reproducibly occurred in plants regenerated from independent replicate cultures and a significant number also overlap with naturally occurring DMRs (Stelpflug et al. 2014). This consistency in the genomic location of DMRs suggests that some loci are susceptible to epigenetic change in response to tissue culture. Greater than 60% of hypomethylated loci were also consistently inherited in self-pollinated progeny plants. By contrast, hypermethylated loci overall were less consistent, less reproducible in independent regenerate cultures, and less heritable. Very similar findings regarding a role for tissue culture in generating DNA methylation have been reported in rice (Stroud et al. 2013).

Inheritance of DNA methylation variation: Genetic variation is highly heritable and exhibits well-known inheritance patterns; however, DNA methylation could be metastable (Regulski et al. 2013). The methylation state of a locus can be influenced by both *cis*- and *trans*-factors (Li et al. 2014b). The combination of these factors raises the possibility of intriguing and unexpected segregation patterns of epialleles. For example, in the case of paramutation, communication of epigenetic state occurs between alleles (Chandler 2007); analysis of inheritance in epiRIL populations also suggests that allelic communication can occur at some, but not all, loci (Johannes et al. 2009; Reinders et al. 2009; Schmitz and Ecker 2012). Similarly, homologous regions located at distant genomic positions can communicate in *trans* as is the case in *PAI* silencing in Arabidopsis (Melquist and Bender 1999). Thus, efforts are ongoing to understand the prevalence and stability of a variety of known and potentially unexpected inheritance patterns.

In general, the methylation state of an allele is faithfully inherited in offspring, whether the parent is selfed or outcrossed. This is also subject to the stochastic changes and reversion that occur over time as noted above. However, both *cis*- and *trans*-factors can influence the methylation state of a locus, including the trans-chromosomal influence of one allele on another. For instance, when alleles with different methylation states are brought together in an F1 hybrid, *trans*-chromosomal methylation (TCM)—a paramutation-like phenomena—can occur, whereby the previously unmethylated loci can become methylated. In turn, this newly methylated state can be inherited in offspring, irrespective of the presence of the original methylated allele, leading to paramutation-like inheritance pattern in F2 plants (Regulski et al. 2013). This is particularly relevant in outcrossing species, such as maize, where there is also significant natural variation in DNA methylation.

Several studies have found that the majority of DMRs are stably inherited in RIL or NIL populations (Eichten et al. 2011; Regulski et al. 2013; Eichten et al. 2013; Li et al. 2014a). In many of these studies, the majority of DMRs investigated were highly stable and exhibited locally inherited (cis) patterns, unaffected by the other allele or other genomic regions. Li et al. (2014b) profiled nearly 1000 DMRs in a large set of NILs and found almost no examples of unstable

inheritance. Only a small number of examples of trans-inheritance were identified, and this investigation did not identify any paramutable loci that displayed consistent characteristics of paramutation across NIL and RIL lines and qPCR validation. In part, experimental design may hamper the identification of trans-acting loci, due to the sequence similarity of interacting loci and requirement for unique alignments during sequencing read mapping in order to profile DNA methylation. Nevertheless, these investigations support the conclusion that the majority of DNA methylation variation in maize is heritable, stable, and mostly controlled in cis.

6.6 Roles of DNA Methylation in Epigenetic Phenomena and Gene Regulation

A primary reason for the interest in DNA methylation is its potential role as a molecular mechanism of epigenetic inheritance. Maize has historically been a model system for the discovery and genetic characterization of epigenetic phenomena including transposable element inactivation, paramutation, and imprinting (Coe 2001). In addition, recent profiles of DNA methylation for multiple inbred lines of maize have revealed substantial natural variation for DNA methylation patterns that might be linked to variation in gene expression. In this section, we will review the evidence for functional roles of DNA methylation in regulating gene expression in epigenetic phenomena and natural variation. Ideally, evidence for functional roles of DNA methylation might be provided through the use of mutant backgrounds or inhibitor treatments that completely abolish DNA methylation. However, there is evidence that severe reductions in DNA methylation in maize are inviable (Li et al. 2014a). Therefore, much of the available evidence for function studies is based on correlative evidence of associations or from studies of plants with minor reductions in methylation at specific contexts (Li et al. 2014a).

6.6.1 Transposable Element Inactivation

Transposable elements (TEs) were first discovered in maize. Very early studies of TEs by McClintock and others documented variation in the activity of these elements, sometimes termed "cycling" or transposable element inactivation (McClintock 1956, 1964). TEs with identical sequence could exist in active or inactive states. Research on maize lines derived from tissue culture found evidence for activation of several DNA transposons coinciding with reduced levels of DNA methylation (reviewed by Kaeppler et al. 2000). These studies provided strong evidence for an association between DNA methylation and transposon activity but did not show that DNA methylation was a required component for silencing TEs. Expression analyses of plants with reductions in CHH (Jia et al. 2009) or CHG methylation (Makarevitch et al. 2007) found evidence for increased transcription of a subset of transposons in the maize genome, but neither study assessed the potential for functional transposon movement. Perhaps the strongest evidence for a functional role of DNA methylation in controlling TE activity is based upon studies of TE activation in maize lines with defective RdDM machinery (Lisch et al. 2002). DNA methylation levels of Mu transposons are reduced in *mop1* plants (Lisch et al. 2002) with defective RDR2 gene (Alleman et al. 2006). Following multiple generations of self pollination in a *mop1* genetic background, there is evidence for stochastic reactivation of Mu elements (Lisch et al. 2002; Woodhouse et al. 2006a, b). These findings may suggest that RdDM activity and CHH methylation is not necessarily required for silencing of *Mu* elements, but is required for stable maintenance of the silenced state (Woodhouse et al. 2006a). Smith et al. (2012) found that treatments of maize tissue cultures with the DNA methylation inhibitor 5-azacytidine could result in reactivation of another transposable element, *TCUP*. This element appears to be regulated by DNA methylation and is often

reactivated during tissue culture (Smith et al. 2012). Studies in Arabidopsis have also provided strong evidence for critical roles of DNA methylation in TE silencing using mutants that affect DNA methylation (reviewed by Underwood et al. 2017). It is likely that DNA methylation is required for the maintained silencing of TEs in the maize genome, and the low viability in genotypes with severe reductions in DNA methylation could be a direct consequence of increased TE expression and transposition.

6.6.2 Paramutation

Paramutation, the directed interaction between two alleles that results in a heritable change in the expression of a paramutable allele following exposure to a paramutagenic allele in a heterozygote, was first discovered at the *r1* (Brink 1956) and *b1* (Coe 1959) loci in maize. Subsequent studies have documented paramutation, or paramutation-like phenomena, at other loci in maize and other species (reviewed by Stam 2009; Hollick 2017). While the genetic sequence of the paramutated locus is the same at the paramutable locus, there is a heritable change in gene expression, providing evidence for epigenetic information. At some paramutated loci, there is evidence for differences in DNA methylation (Eggleston et al. 1995; Walker 1998; Sidorenko and Peterson 2001) or other chromatin marks (Haring et al. 2010). However, the exact nature of molecular mechanisms involved in establishing and maintaining paramutated states remain unclear. Genetics screens have uncovered a number of factors required for paramutation (reviewed by Hollick 2017), including components of the RdDM pathway as well as other chromatin genes, providing evidence that RdDM and/or DNA methylation is necessary for maintenance of the paramutated epigenetic state at some loci (Alleman et al. 2006; Hale et al. 2007; Barber et al. 2012). The fact that multiple components of the RdDM pathway have been isolated through forward genetic screens to find factors involved in paramutation certainly suggests a functional linkage between DNA methylation and paramutation. However, it is worth noting that only components of the RdDM pathway, not pathways involved in maintenance of CG or CHG methylation, have been recovered. This could indicate a special role for CHH methylation or could suggest that the siRNAs produced and utilized by RdDM are critical for paramutation. Alternatively, this could be due to the fact that severe reductions in CG or CHG methylation may be inviable.

6.6.3 Imprinting

Imprinting (reviewed by Gehring 2013) was first characterized in maize based on differential expression of the transcription factor from the *R* locus depending upon whether this locus was inherited from the maternal or paternal parent (Kermicle 1970). Similar patterns upon parent-of-origin-dependent seed color can also be observed for some alleles of the *B* locus (Selinger et al. 2001). Recent genome-wide surveys of imprinting in the maize endosperm have revealed several hundred imprinted genes in maize (Zhang et al. 2011; Waters et al. 2011). Differential methylation of the maternal and paternal alleles has been documented for several of the well-characterized imprinted genes (Haun et al. 2007; Hermon et al. 2007). Lauria et al. (2004) documented evidence for extensive hypomethylation of the maternal genome in maize endosperm tissue. Based on studies in Arabidopsis and rice, where a similar phenomenon is found (Jullien et al. 2012), it is likely that this is due to expression of the DNA demethylase enzyme DME in the central cell prior to fertilization (Park et al. 2016). This global reduction of DNA methylation is then maintained following fertilization and results in reduced methylation of the maternal alleles at some loci in endosperm tissue in Arabidopsis and rice (reviewed by Gehring 2013). A genome-wide analysis of DNA methylation in the maize endosperm reveals thousands of parent-of-origin DMRs (pDMRs) with many of these located near genes with imprinted expression patterns (Zhang et al. 2014). There is also evidence for histone

modification differences, particularly H3K27me3, between the maternal and paternal alleles at numerous imprinted loci that may be more important for imprinting than DNA methylation (Haun and Springer 2008; Zhang et al. 2014). Interestingly, reduced DNA methylation of the maternal allele can be associated with both maternally expressed genes (MEGs) and paternally expressed genes (PEGs) suggesting that the DNA methylation is not necessarily required for silencing during imprinting. Indeed, PEGs are more often associated with DNA methylation than MEGs (Gehring et al. 2011). In these cases, the hypomethylated maternal allele often is associated with high levels of H3K27me3 and reduced methylation may be required to allow for this other silencing mark to be added (Wolff et al. 2011; Makarevitch et al. 2013). There are also many imprinted genes that do not contain evidence for altered methylation of the maternal and paternal alleles (Waters et al. 2011; Zhang et al. 2014), suggesting that not all examples of imprinting require allelic DNA methylation differences.

6.6.4 DNA Methylation and Natural Variation for Gene Expression

DNA methylation could also play a critical role in generating epialleles, differences in gene expression without changes in DNA sequence. DNA methylation profiling has revealed abundant examples of natural variation for DNA methylation (DMRs) (Eichten et al. 2011; Regulski et al. 2013; Eichten et al. 2013; Li et al. 2015b). In several cases, RNAseq and DNA methylation data has been collected in matched tissue samples providing an opportunity to assess potential associations between DNA methylation and gene expression levels. Eichten et al. (2013) assessed the connection between DNA methylation and gene expression for 1,966 DMRs present in multiple inbred lines and located within 10 kb of a maize gene, and 277 examples of a significant association were documented (Eichten et al. 2013). The majority of cases reflect a negative association in which increased DNA methylation is associated with reduced or absent gene expression. Whole-genome bisulfite sequencing of 5 maize inbreds identified a large number of context-specific DMRs in maize (Li et al. 2015b). RNAseq data on the same tissues was used to identify differentially expressed genes. A comparison of DNA methylation levels in the region surrounding the transcription start site revealed that genes with moderate changes in gene expression (fivefold change or less) are not enriched for DMRs. However, genes with tenfold or greater changes in gene expression are enriched for DMRs in the promoter region. Approximately 20% of genes that exhibit qualitative (on–off) differences in expression exhibit altered methylation in regions surrounding the transcription start site (Li et al. 2015b). In combination, these two studies provide evidence that DNA methylation changes are associated with some examples of natural variation for gene expression in maize and are more often found at genes with qualitative variation in expression. Makarevitch et al. (2007) provided more direct evidence for a role of DNA methylation in natural variation for gene expression in maize. The *zmet2-m1* mutation, which results in reduced CHG methylation (Papa 2001; Li et al. 2014a), was introgressed into multiple genetic backgrounds, and these stocks were used for expression profiling. Interestingly, the genes that are up-regulated in the *zmet2-m1* mutant lines relative to wild-type controls were significantly different in B73, Mo17, and W22. Many of these genes are expressed in wild type of some lines but silent in the others and loss of CHG methylation in the mutant results in activation of these genes. There is also evidence that natural variation for DNA methylation may result in variation in splicing patterns among different inbred lines (Regulski et al. 2013; Mei et al. 2017).

6.7 Concluding Remarks

The epigenome has the potential to provide additional heritable information that can influence traits in maize and other plant species. Our

ability to document the genome-wide distribution of DNA methylation in maize has provided clues to the potential for this information to influence gene expression and plant traits. Such analysis has also revealed important distinctions between Arabidopsis and maize. Continued research will be necessary to better understand the molecular mechanisms that control DNA methylation in maize and to elucidate the sources of variation for DNA methylation. It will be important to document whether substantial levels of variation in DNA methylation are uncoupled from nearby SNPs because these will not be captured in SNP-based selection schemes. We anticipate exciting advances in our understanding of the functional relevance of DNA methylation and other chromatin modifications in maize in the coming years.

References

Alleman M et al (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442(7100):295–298

Anderson SN et al (2018) Subtle perturbations of the maize methylome reveal genes and transposons silenced by chromomethylase or RNA-directed DNA methylation pathways

Barber WT et al (2012) Repeat associated small RNAs vary among parents and following hybridization in maize. Proc Natl Acad Sci USA 109(26):10444–10449

Baucom RS et al (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet 5(11):e1000732

Bauer MJ, Fischer RL (2011) Genome demethylation and imprinting in the endosperm. Curr Opin Plant Biol 14 (2):162–167

Becker C et al (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480 (7376):245–249

Bewick AJ et al (2016) On the origin and evolutionary consequences of gene body DNA methylation. Proc Natl Acad Sci USA 113(32):9111–9116

Brink RA (1956) A genetic change associated with the R locus in maize which is directed and potentially reversible. Genetics 41(6):872–889

Chandler VL (2007) Paramutation: from maize to mice. Cell 128(4):641–645

Coe EH (1959) A regular and continuing conversion-type phenomenon at the B locus in maize. Proc Natl Acad Sci USA 45(6):828–832

Coe EH (2001) The origins of maize genetics. Nat Rev Genet 2(11):898–905

Crisp PA et al (2016) Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics. Sci Adv 2(2):e1501340

Du J et al (2014) Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. Mol Cell 55(3):495–504

Du J et al (2015) DNA methylation pathways and their crosstalk with histone methylation. 33(4):395–401

Du J et al (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151(1):167–180

Eggleston WB, Alleman M, Kermicle JL (1995) Molecular organization and germinal instability of R-stippled maize. Genetics 141(1):347–360

Eichten SR et al (2011) Heritable epigenetic variation among maize inbreds. PLoS Genet 7(11)

Eichten SR et al (2013) Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell 25(8):2783–2797

Eichten SR et al (2012) Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS Genet 8(12)

Eichten SR, Springer NM (2015) Minimal evidence for consistent changes in maize DNA methylation patterns following environmental stress. Front Plant Sci 6:308

Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. Science 330(6004):622–627

Forestan C, Farinati S, Aiese Cigliano R, Lunardon A, Sanseverino W, Varotto S (2017) Maize RNA PolIV affects the expression of genes with nearby TE insertions and has a genome-wide repressive impact on transcription. BMC Plant Biol 17:161

Gehring M (2013) Genomic imprinting: insights from plants. Annu Rev Genet 47:187–208

Gehring M, Missirian V, Henikoff S (2011) Genomic analysis of parent-of-origin allelic expression in arabidopsis thaliana seeds. PLoS ONE 6(8):e23687

Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324 (5933):1447–1451

Gent JI et al (2013) CHH islands: De novo DNA methylation in near-gene chromatin regulation in maize. Genome Res 23(4):628–637

Gouil Q, Baulcombe DC (2016) DNA methylation signatures of the plant chromomethyltransferases. PLoS Genet 12(12):e1006526

Greenberg MVC et al (2013) Interplay between active chromatin marks and RNA-directed DNA methylation in arabidopsis thaliana. PLoS Genet 9(11)

Haag JR et al (2014) Functional diversification of maize RNA polymerase IV and V subtypes via alternative catalytic subunits. 9(1), 378–390

Hale CJ et al (2007) A novel Snf2 protein maintains trans-generational regulatory states established by paramutation in maize. PLoS Biol 5(10):2156–2165

Haring M et al (2010) The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. Plant J Cell Mol Biol 63(3):366–378

Haun WJ et al (2007) Genomic imprinting, methylation and molecular evolution of maize enhancer of zeste (Mez) homologs. Plant J Cell Mol Biol 49(2):325–337

Haun WJ, Springer NM (2008) Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J Cell Mol Biol

Heard E, Martienssen RA (2014) Transgenerational epigenetic inheritance: myths and mechanisms. Cell 157(1):95–109

Henderson IR et al (2010) The De novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in arabidopsis thaliana. PLoS Genet 6(10):1–11

Hermon P et al (2007) Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. Plant Mol Biol 64(4):387–395

Hollick JB (2017) Paramutation and related phenomena in diverse species. Nat Rev Genet 18(1):5–23

Hsieh T-F et al (2009) Genome-wide demethylation of arabidopsis endosperm. Science 324(5933):1451–1454

Jia Y et al (2009) Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS Genet 5 (11):e1000737

Jiao Y et al (2017) Improved maize reference genome with single-molecule technologies. Nature 546 (7659):524–527

Johannes F et al (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genet 5(6):e1000530

Johnson LM et al (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature 507(7490):124–128

Jullien PE et al (2012) DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr Biol CB 22(19):1825–1830

Kaeppler SM, Phillips RL (1993) Tissue culture-induced DNA methylation variation in maize. Proc Natl Acad Sci USA 90(19):8773–8776

Kaeppler SM, Kaeppler HF, Rhee Y (2000) Epigenetic aspects of somaclonal variation in plants. Plant Mol Biol 43(2–3):179–188

Kawakatsu T et al (2016) Unique cell-type-specific patterns of DNA methylation in the root meristem. Nat Plants 2:16058

Kermicle JL (1970) Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics 66(1):69–85

Lauria M et al (2004) Extensive maternal DNA hypomethylation in the endosperm of Zea mays. Plant Cell 16(2):510–522

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204–220

Law JA et al (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498(7454):385–389

Li Q et al (2014a) Genetic perturbation of the maize methylome. Plant Cell 26(12):4602–4616

Li Q et al (2014b) Inheritance patterns and stability of DNA methylation variation in maize near-isogenic lines. Genetics 196(3):667–676

Li Q et al (2015a) RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. In: Proceedings of the National Academy of Sciences of the United States of America, p 1514680112

Li Q et al (2015b) Examining the causes and consequences of context-specific differential DNA methylation in maize. Plant Physiol 168(4):1262–1274

Li Q et al (2015c) Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. Nucleic Acids Res 43(12):1–16

Lisch D et al (2002) A mutation that prevents paramutation in maize also reverses mutator transposon methylation and silencing. Proc Natl Acad Sci USA 99(9):6130–6135

Lister R et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133(3):523–536

Liu R et al (2015) A DEMETER-like DNA demethylase governs tomato fruit ripening. Proc Natl Acad Sci USA 112(34):10804–10809

Loenen WAM, Raleigh EA (2014) The other face of restriction: modification-dependent enzymes. Nucleic Acids Res 42(1):56–69

Lu Y, Rong T, Cao M (2008) Analysis of DNA methylation in different maize tissues. J Genet Genomics 35(1):41–48

Makarevitch I et al (2013) Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. Plant Cell 25(3):780–793

Makarevitch I et al (2007) Natural variation for alleles under epigenetic control by the maize chromomethylase Zmet2. Genetics 177(2):749–760

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation : an epigenetic pathway of increasing complexity

McClintock B (1956) Controlling elements and the gene. Cold Spring Harb Symp Quant Biol 21:197–216

McClintock B (1964) Aspects of gene regulation in maize. Carnegie Inst Wash Year Book 63:592–602

McCue AD et al (2014) ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the

establishment of DNA methylation. EMBO J 34 (1):20–35
Mei W et al (2017) A comprehensive analysis of alternative splicing in paleopolyploid maize. Front Plant Sci 8:1–19
Melquist S, Luff B, Bender J (1999) Arabidopsis PAI gene arrangements, cytosine methylation and expression. Genetics 153(1):401–413
Niederhuth CE et al (2016) Widespread natural variation of DNA methylation within angiosperms. bioRxiv 1–19
Palmer LE (2003) Maize genome sequencing by methylation filtration. Science 302(5653):2115–2117
Panda K, Slotkin RK (2013) Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. Plant Signal Behav 8(8):8–10
Papa CM (2001) Maize chromomethylase zea methyltransferase2 is required for CpNpG methylation. Plant Cell Online 13(8):1919–1928
Park M, Keung AJ, Khalil AS (2016) The epigenome: the next substrate for engineering. Genome Biol 17 (1):183–185
Pecinka A, Scheid OM (2012) Stress-induced chromatin changes: a critical view on their heritability. Plant Cell Physiol p.pcs 044
Phillips RL, Kaeppler SM, Olhoft P (1994) Genetic instability of plant tissue cultures: breakdown of normal controls. Proc Natl Acad Sci USA 91 (12):5222–5226
Rabinowicz PD et al (2005) Differential methylation of genes and repeats in land plants. Genome Res 15 (10):1431–1440
Regulski M et al (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res 23(10):1651–1662
Reinders J et al (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes Dev 23(8):939–950
Richards EJ (2006) Inherited epigenetic variation—revisiting soft inheritance. Nat Rev Genet 7(5):395–401
Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends Plant Sci 17:149–154
Schmitz RJ et al (2011) Transgenerational epigenetic instability is a source of novel methylation variants. Science 334(6054):369–373
Schnable PS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. 326
Secco D et al (2015) Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. eLife 4. https://doi.org/10.7554/elife.09343
Selinger DA, Chandler VL (2001) B-Bolivia, an allele of the maize b1 gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. Plant Physiol 125(3):1363–1379
Sidorenko LV, Peterson T (2001) Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. Plant Cell 13(2):319–335
Smith AM, Hansey CN, Kaeppler SM (2012) TCUP: a novel hAT transposon active in maize tissue culture. Front Plant Sci 3:6
Springer NM, Schmitz RJ (2017) Exploiting induced and natural epigenetic variation for crop improvement. Nat Rev Genet. In press
Stam M (2009) Paramutation: a heritable change in gene expression by allelic interactions in trans. Mol Plant 2 (4):578–588
Stelpflug SC et al (2014) Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. Genetics 198(1):209–218
Stonaker JL et al (2009) Diversity of Pol IV function is defined by mutations at the maize rmr7 locus. PLoS Genet 5(11)
Stroud H et al (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21(1):64–72
Stroud H et al (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152(1–2):352–364
The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408(6814):796–815
To TK, Saze H, Kakutani T (2015) DNA methylation within transcribed regions. Plant Physiol 4(1):00543
Underwood CJ, Henderson IR, Martienssen RA (2017) Genetic and epigenetic variation of transposable elements in Arabidopsis. Curr Opin Plant Biol 36:135–141
van der Graaf A et al (2015) Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. Proc Natl Acad Sci USA 112(21):6676–6681
Walker EL (1998) Paramutation of the r1 locus of maize is associated with increased cytosine methylation. Genetics 148(4):1973–1981
Wang P et al (2015) Genome-wide high-resolution mapping of DNA methylation identifies epigenetic variation across embryo and endosperm in Maize (Zea may). BMC Genom 16:21
Waters AJ et al (2011) Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell 23(12):4221–4233
West PT et al (2014) Genomic distribution of H3K9me2 and DNA methylation in a maize genome. PLoS ONE 9(8):1–10
Wolff P et al (2011) High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. PLoS Genet 7(6):e1002126
Woo HR, Dittmer TA, Richards EJ (2008) Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in arabidopsis. PLoS Genet 4(8)
Woodhouse MR, Freeling M, Lisch D (2006a) Initiation, establishment, and maintenance of heritable MuDR transposon silencing in maize are mediated by distinct factors. PLoS Biol 4(10):e339

Woodhouse MR, Freeling M, Lisch D (2006b) The mop1 (mediator of paramutation1) mutant progressively reactivates one of the two genes encoded by the MuDR transposon in maize. Genetics 172(1):579–592

Zemach A et al (2013) The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153 (1):193–205

Zemach A et al (2010) Local DNA hypomethylation activates genes in rice endosperm. Proc Natl Acad Sci USA 107(43):18729–18734

Zhang H, Zhu JK (2012) Active DNA demethylation in plants and animals. Cold Spring Harb Symp Quant Biol 77:161–173

Zhang M et al (2014) Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Res 24(1):167–176

Zhang M et al (2011) Tissue-specific differences in cytosine methylation and their association with differential gene expression in sorghum. Plant Physiol 156 (4):1955–1966

Zilberman D et al (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet 39(1):61–69

Zong W et al (2013) Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress. Plant Mol Biol 81(1–2):175–188

7 Integrating Transcriptome and Chromatin Landscapes for Deciphering the Epigenetic Regulation of Drought Response in Maize

Cristian Forestan, Silvia Farinati, Alice Lunardon and Serena Varotto

Abstract

Water scarcity associated with climate change is among the principal constraints to plant productivity worldwide, and crop growth models predict that this issue will be more severe in future. Plants withstand drought stress by modifying their gene expression patterns and activating a variety of physiological and biochemical responses at cellular and whole-organism levels. Molecular and genomic studies have indeed identified many stress-inducible genes in different plant species. Stress-responsive genes encode for proteins with various functions and signaling factors, such as transcription factors, protein kinases, and protein phosphatases, but also include several noncoding and regulatory RNAs involved in the modulation of the stress response networks, making it a very complex phenomenon. Affecting a number of different aspects of plant growth and development, chromatin-based mechanisms, such as histone post-translational modifications, are fundamental for the fine coordination and tuning of gene expression in response to environmental cues. Several histone modifications have been found dramatically altered on stress-responsive gene regions under drought stress; thus, the integration of different omics technologies are essential to deeply understand plant tolerance mechanisms and manage them toward breeding for drought tolerance in maize.

7.1 Introduction

Although agriculture has been facing drought since ancient time, climate change is evidently increasing the frequency of this phenomenon and severity in some areas of the world, causing significant yield reductions in major cereal species including maize (Abdul Jaleel et al. 2009). Keeping in mind that the correlation between climate change and global-scale drought trends is a matter of debate because of the difficulties in distinguishing decadal-scale variability from long-term trends, climate change is certainly making rainfall patterns less predictable: in some areas of the world, the precipitations are going to increase, but in some others, drought will cause significant issues. Model projections forecast an increase in dry areas in both Mediterranean area and southwest North America (Cook et al. 2016;

C. Forestan · S. Farinati · S. Varotto (✉)
Department of Agronomy, Food, Natural resources, Animal and Environment (DAFNAE) Agripolis, University of Padova, Viale dell'Università 16, 35020 Legnaro, PD, Italy
e-mail: serena.varotto@unipd.it

A. Lunardon
Department of Biology and Huck Institutes of the Life Sciences, Penn State University, University Park, Los Angeles, PA 16802, USA

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_7

Seager et al. 2009). Rural areas are expected to struggle major impacts on water availability, and competition for water resources among different social and economic sectors is also expected to grow, with agriculture being progressively forced to use low-quality water (Laraus 2004). Moreover, estimation on crop yields in tropical and temperate regions is projected to further decline if predicted local temperature increases of 2 °C above those seen in the late twentieth century are observed. In this scenario, it is expected that sensitive crops, like maize, will be particularly damaged by climate change.

Because maize has a pronounced susceptibility to drought (Banziger and Araus 2007), it is important to develop new maize varieties that are more adaptable and tolerant to changes in water availability, so that they can maintain optimum yield levels under water stress conditions (Harrison et al. 2014; Lobell et al. 2014). To address any future constraints related to maize culturing in areas with water limitations, there is the need to better characterize relationships between water stress and maize productivity. Genetic and epigenetic studies of traits for tolerance/susceptibility to drought will improve our knowledge on adaptation to changing climatic conditions. Elucidating the mechanisms through which maize responds to water stress and manipulating these mechanisms is fundamental to enhance maize productivity in suboptimal environments.

To survive in a changing environment, plants have elaborated adaptive strategies largely based on changes in gene expression. These changes are mediated by signaling cascades that plants use for continuously monitoring external signals, in order to align development with environment. Stress-induced changes in metabolism and development are linked to altered pathways in gene expression that start from a single cell and are transmitted throughout the whole plant. The scale and timing of stress response is dictated by the severity, the timing, and the duration of the stress. Epigenetic mechanisms, such as DNA methylation, chromatin modifications, and small RNAs, have a fundamental role in spatiotemporal gene expression changes during stress response and adaptation (Asensi-Fabado et al. 2017). Investigating gene product accumulation patterns, gene functions, and gene expression regulation mechanisms is therefore pivotal to understand the plants' stress response.

Global transcriptome analysis, including the noncoding portion of the genome, and chromatin immunoprecipitation (ChIP) are major sources of information used to unravel gene networks and chromatin histone modifications landscapes. In this chapter, we will present how these approaches can be applied to advance our understanding of maize genetic and epigenetic responses to water stress. Coupling next-generation sequencing (NGS) techniques with transcriptome and chromatin analyses provides genome-wide data to evaluate maize lines under water-deficit conditions, to better understand how maize plants modulate their drought response, as well as during the recovery from stress. Integrating omics data allows a better understanding on how epigenetic mechanisms work as an intermediary between the environment and the genome.

7.2 Analysis of Transcriptome Profiles During Drought Stress Response in Maize

A fair amount of genetic studies have focused on the controlling mechanisms for plant performance under drought in maize and other crop species, and most of them have shown that drought stress tolerance is a complex trait (Haak et al. 2017; Miao et al. 2017; Shinozaki and Yamaguchi-Shinozaki 2007) that involves stress signal perception, transduction, and expression of downstream regulatory and functional effector genes. The understanding of the interplay between genetic variegate components, comprising coding genes and their regulatory network, requires whole-genome transcriptome profiling for the identification of transcription factors, coding gene families, and noncoding RNA regulatory components involved in drought tolerance. The genome-wide measurement of transcript expression levels in different plant growth conditions provides the ability to quantify

the modulation in gene expression following stress applications. RNA sequencing (RNA-Seq) is the most recent approach in transcriptome analysis; it does not need a reference genome and allows entire transcriptomes to be surveyed in a high-throughput and quantitative manner (Conesa et al. 2016). The efforts that have been made in candidate gene discovery through RNA deep sequencing have demonstrated that many of the pathways important for plant growth under limited water tend to be conserved among plant species. The involved pathways comprise transcription regulation, signal transduction, protein biosynthesis and decay, membrane trafficking, and photosynthesis (Nakashima et al. 2014). However, we have to point out that conservation of genes and pathway is not sufficient to translate results from one species to another because the high conservation of the core gene machinery between plants may not correlate with the expression timing of stress-induced genes.

Sequencing of plant total RNA, in addition to mRNA abundance of protein-coding genes, can also lead to the sequencing of noncoding RNAs (ncRNAs), which are functional RNAs with low protein-coding potential (Forestan et al. 2016; Xu et al. 2017; Zhang et al. 2014; Zhang and Chen 2013), and transposable element (TE) transcripts, and more specifically of expressed retrotransposon that can move around the genome via RNA intermediates (Forestan et al. 2016; Wang et al. 2017a). ncRNAs can be divided into small ncRNAs (sRNAs; 18–30 nt) and long ncRNAs (lncRNAs; >200 nt). Plant micro-RNAs (miRNAs) are commonly 21-nt sRNAs and guide degradation and/or translation inhibition of their mRNA targets in the so-called post-transcriptional gene silencing (PTGS) process (Axtell 2013; Kumar et al. 2018). Small interfering 21–24-nt RNAs (siRNAs) can suppress the presence of their target genes in the cytoplasm and also activate transcriptional gene silencing (TGS) mechanisms by establishing chromatin modifications in the nucleus (Matzke and Mosher 2014). Sequencing these regulatory noncoding RNAs and TEs is important because many noncoding RNAs are modulated by stresses (Liu et al. 2015; Wang et al. 2017b) and could function as *cis*- or *trans*-regulators of gene expression, and because various stresses have been reported to activate TEs transcription and transposition in plants. This is particularly true in maize, which is an ancient allotetraploid species characterized by a highly repetitive genome (>85% TE sequences; Baucom et al. 2009).

7.2.1 Setting-up and Validation of an Agronomical-Realistic Drought Stress Protocol

Environmental stresses are three-dimensional entities defined by their timing, duration, and severity (Blum 2014). In published studies aimed to identify maize drought-responsive genes, osmotic stress has been mostly simulated in vitro by the addition of polyethylene glycol (PEG) or, when the plants have been grown in soils, they have been sampled soon after germination or at seedling stages after hours or a few days of drought treatment (Jia et al. 2006; Lu et al. 2011; Opitz et al. 2014; Shan et al. 2013; Wu et al. 2017). To study maize stress response at genetic and epigenetic genome-wide levels, it is fundamental to set reproducible drought stress protocols, with time-limited stress conditions assessable at agronomical and physiological levels that retain their value under field conditions and can be useful for agronomical solutions. For example, in field conditions, limited availability of resources rarely causes plant death. Instead, after a period of stress, environmental conditions usually turn more favorable, compromising the crop yield but not the plant survival (Morari et al. 2015).

Prior to starting a gene expression profiling study of maize drought stress response, we developed a preliminary stress protocol that mimics field progressive stress conditions and evaluated the stress response at the physiological level in the time course of application (Morari et al. 2015). The experiment was conducted at the experimental farm of the University of Padova, Italy, during spring–summer in a field with an automatic mobile roof for avoiding rainfall input. Maize plants were grown in pots

that were weighted and watered daily. Unstressed plants were grown at a water content of 100% available water capacity, replenishing every day the water lost by evapotranspiration. Water-stressed plants were irrigated replenishing only 60% of daily evapotranspiration. Initially, stress response at the physiological level was evaluated in two maize genotypes: the reference inbred line B73 and a F1 hybrid selected for its tolerance to stress. The stress treatment started at the vegetative 6 (V6) stage, and plants were sampled at the beginning of treatments (T0) after four days (T4), and after ten days (T10) days, plus non-treated respective controls. For evaluating plant recovery capacity after T10, all plants were watered twice a day. This pilot experiment allowed us to determine that when considering both plant growth and physiological data on net assimilation, stomatal conductance and quantum efficiency of photosystem II, the applied stress was effective in limiting both root and shoot growth in the hybrid and arresting the growth in the inbred line. In addition, the experiment results indicated that B73 shoots needed a longer recovery time than 4 days (T14) to start growing again and reaching a new homeostasis after the stress application. During this initial experiment and for setting an effective protocol for analyzing the stress effects at a genome-wide level, the transcript level of genes previously showed to be markers of drought condition and belonging to different stress response pathways was determined using quantitative PCR. Some of the analyzed stress markers were affected both by the stress duration and severity, confirming previous literature data in our specific long drought stress conditions. Interestingly, these effects varied between genotypes and they showed a high correlation with the stress response at the physiological level (Morari et al. 2015).

Using the results produced by this preliminary experiment and with the aim to study the stress response at genetic and epigenetic levels using a genome-wide approach, a second stress experiment was set using the B73 inbred line and the *Required to Maintain Repression 6* mutant (*rmr6/rpd1-1*, involved in siRNA biogenesis and in the RNA directed DNA methylation pathway —Erhard et al. 2009, 2015—introgressed in the B73 genetic background). The osmotic stresses (drought, discussed in detail below, salinity and drought plus salinity, not further discussed for the purpose of this chapter) were applied for ten days (T10) followed by a seven-day recovery period (T17) plus controls for each genotype and treatment. Plants were regularly watered till pot capacity until the V5/V6 developmental stage, when stress treatments were applied: control plants (NS) were watered with 75% of disposable water and drought-stressed plants (WS) with 25% of disposable water. In the tenth day of treatment and in the seventh day of recovery, the youngest wrapped leaf was harvested from each plant sample (Forestan et al. 2016; Lunardon et al. 2016). Leaves of the same genotype, treatment, and sampling time points were pooled together and flash-frozen in liquid nitrogen and stored at −80 °C. The complete experiment was replicated three times, and leaf samples were used to perform RNA-Seq, sRNA-Seq, and ChIP-Seq, to obtain a global view of the molecular response to drought stress, including the noncoding portion of genome, and to dissect the characteristics of the stress-recovery response, by investigating whether a transitory stress can cause a sort of memory of the stress.

7.2.2 RNA-Seq Analysis of Drought-Induced Transcriptional Changes in Maize Leaves

The power RNA-Seq lies in the opportunity to combine transcript discovery and quantification in a single assay, capturing the most complete picture of the sample transcriptome, including coding and multiple forms of noncoding RNAs. In order to improve the maize predicted gene models and analyze their expression under osmotic stress conditions, an extensive set of RNA-Seq data covering the stress leaf transcriptome of maize were produced in our laboratory (Forestan et al. 2016).

Total RNA was extracted from developing leaves of maize plants subjected to the three

osmotic stresses and recovery stages and used for directional Illumina sequencing. A total RNA-Seq assay that relies on ribosomal RNA removal from extracted RNA instead of enrichment for mRNA using poly(A) selection was applied. After sequencing and following the quality trimming and filtering steps, high-quality reads were aligned to the maize reference genome (RefGen ZmB73 Assembly AGPv3 and Zea_mays.AGPv3.20.gtf transcript annotation) and used for a de novo assembly approach to generate the complete transcriptome of the maize leaf in response to abiotic stresses. The "Reference Annotation Based Transcript" (RABT) assembly method builds upon the Cufflinks assembler (Trapnell et al. 2012), allowed the identification of more than 25,000 new maize transcripts, primarily accounting for novel splicing variants at known loci, but also for newly identified intergenic transcribed loci and antisense transcripts (NATs) mapping with opposite orientation in respect to reference annotated transcripts (Forestan et al. 2016).

The newly annotated transcripts were functionally characterized firstly by means of gene ontology (GO) annotation and then by the evaluation of their coding potential. Specific GO terms were assigned to more than 80% of new isoforms and to only 9% of intergenic and 4% of antisense transcripts, indicating a low protein-coding potential of these two classes. GO enrichment analysis revealed that new splicing variants are enriched for categories related to mitosis/cell cycle, gene expression regulation, RNA-mediated gene silencing, chromosome organization, and protein modification. Over-representation for DNA recombination/integration, protoderm and meristem differentiation terms characterize intergenic transcripts, for the 9% of intergenic transcripts that could be functionally annotated. Finally, ncRNA metabolism and plant embryonic and post-embryonic development-associated GO terms were enriched among natural antisense transcripts (NATs), which were prevalently identified at the opposite strand of annotated transcription factors.

The whole maize transcriptome was further functionally characterized by the application of a computational pipeline that allowed the systematic identification of 13,387 long noncoding RNAs (lncRNAs) and 21,624 putative precursors of sRNAs, that may play critical roles in regulating gene expression through multiple RNA-mediated mechanisms (Forestan et al. 2016).

Focusing on B73 drought-stressed and recovery samples, pairwise differential expression analyses revealed hundreds of differentially expressed genes (DEG; fold change in expression $\geq |2|$ and FDR—adjusted p value ≤ 0.05) in response to long-term stress application, comprising lncRNAs and transposable elements (Fig. 7.1a). After 10 days of progressive drought stress (WST10), 797 up- and 430 down-regulated gene compared to control sample (NST10), were identified. Of these, 80% are represented by protein-coding genes, while the residual part included ncRNAs (both lncRNA and sRNA precursors) and transposable element-related loci. When comparing stressed and recovered plants, 155 and 169 genes were identified as up- and down-regulated, respectively, from WST10 to WST17: about 80% of the 169 down-regulated genes were previously identified as up-regulated in WST10, indicating a transient, stress-induced transcriptional change at these loci. On the contrary, only 40% of the 155 WST17 up-regulated genes were misregulated in WST10 versus NST10, suggesting their specific involvement in the stress-recovery molecular regulation. Finally, after the recovery period, plants that experienced the stress (WST17) significantly misregulated 41 genes compared to control plants at the same time point (NST17) indicating that the drought stress could affect transcription after a long recovery period. Interestingly, ncRNAs represent a great proportion of misregulated genes: this observation suggests that they might be directly involved in gene regulation during the drought stress response and adaptation.

ncRNAs stress-responsive expression and their key regulatory role in stress response and tolerance were demonstrated in different plant species, comprising *Arabidopsis thaliana* (Di et al. 2014; Matsui et al. 2013; Qin et al. 2017) and crops such as rice (Chung et al. 2016), maize

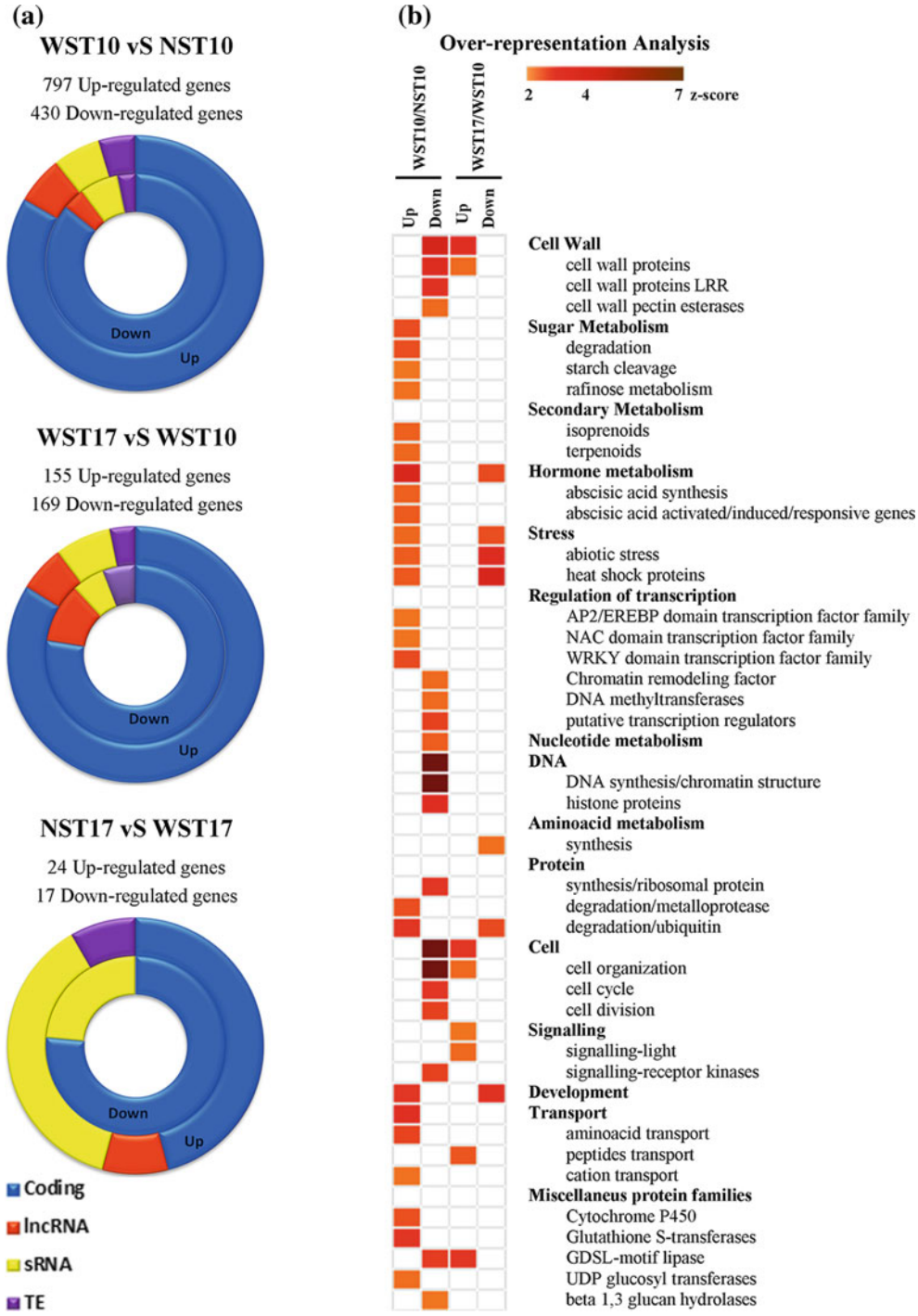

Fig. 7.1 Summary and functional annotation of drought differentially expressed genes. **a** Summary of genes differentially expressed in each pairwise comparison divided by their coding/no coding classification. In the ring graphs, the inner ring depicts the down-regulated genes and the outer the up-regulated ones. **b** MapMan functional categories enriched among genes DE in each pairwise comparison. Over-representation of categories was determined using Fisher's exact test, and resulting p values were adjusted according to Benjamini and Hochberg. Z scores automatically calculated from p values (e.g., 1.96 corresponds to a P value of 0.05) are plotted in an orange to brown color scale

(Zhang et al. 2014), and sugarcane (Lembke et al. 2012), as has been recently reviewed in Wang et al. (2017b). Furthermore, a recent work confirmed the high impact of water stress in natural antisense transcripts accumulation in maize and their involvement in both transcriptional and post-transcriptional gene regulation (Xu et al. 2017), being NATs also associated with the production of regulatory sRNAs. Lastly, a contribution of transposable elements to the regulation of maize genes in response to abiotic stresses (Makarevitch et al. 2015) was recently demonstrated as well, corroborating the existence of very complex transcriptional network activation during the response and adaptation to environmental challenges in this species.

To functionally characterize the drought differentially expressed genes, GO enrichment analysis (determined using a custom maize GO annotation including the newly identified transcripts; Forestan et al. 2016) was coupled with MapMan category over-representation (Thimm et al. 2004; Usadel et al. 2009). Genes up-regulated in WST10 compared to NST10 display obvious enrichment for GO categories related to "stress response," "lipid metabolism," "abscisic acid-signaling," and "cell-death." Stress down-regulated genes showed instead a strong enrichment in many functional terms, including cell differentiation, plant growth, development, and morphogenesis, cell-wall, plasma membrane, cell cycle regulation of gene expression, and epigenetic regulation-related categories. Similar results were obtained with the MapMan software, which also highlighted the over-representation of several transcription factor families among the stress up-regulated genes (Fig. 7.1b).

AP2/EREBP, NAC, and WRKY represent well-known transcription factor families responsive to drought (Chen and Zhu 2004; Joshi et al. 2016). *AP2/EREBP* (APETALA2/ethylene-responsive element binding proteins) transcription factors form a large multigene family unique to plants and represent key regulators of numerous plant growth processes, from cell identity determination to response to various types of biotic and environmental stresses (Riechmann and Meyerowitz 1998). They regulate the expression of stress-related genes in an ABA-independent manner, and their over-expression in transgenic plants has been shown to enhance tolerance toward osmotic stress in rice (Cui et al. 2011; Mizoi et al. 2012; Paul et al. 2015), potato (Iwaki et al. 2013), sugarcane (Augustine et al. 2015), and wheat (Shavrukov et al. 2016). Recently, it was also demonstrated that osmotic stress activates the transcription of the maize AP2/EREBP family member *ZmDREB2A* by acting on chromatin accessibility through histone acetylation within its promoter region (Zhao et al. 2014). For all these reasons, *AP2/EREBP* transcription factors are considered as important candidates for stress tolerance engineering (Agarwal et al. 2017; Dey and Corina Vlot 2015; Liu et al. 2013).

Also *NAC* transcription factors are part of a large gene family and are involved in the regulation of drought-related genes as transcriptional activators or repressors (Joshi et al. 2016; Wang et al. 2016). In rice, several *NAC* genes are induced during early stages of salt and drought stresses (Hong et al. 2016) and they confer drought tolerance through induction of downstream *PP2C* genes (You et al. 2014), target genes that resulted as highly over-expressed also in WST10 stressed leaf samples. Another study reported that rice *NAC1* was induced under drought stress in guard cells, increasing stomatal closure to prevent transpirational water losses (Singh et al. 2015). Interestingly, a transposable element insertion in the promoter region of maize *NAC111* was significantly associated with natural variation in maize drought tolerance (Mao et al. 2015).

Lastly, WRKY transcription factors are key components of signaling networks that modulate many plant processes and are known to function also in adaptation to biotic and abiotic stresses (Okay et al. 2014; Phukan et al. 2016). For example, over-expression of rice *WRKY11*, *WRKY45,* and *WRKY72* results in enhanced drought tolerance (Ding et al. 2014; Qiu and Yu 2009; Wu et al. 2009; Yu et al. 2010), but other WRKY TFs are involved in Pi homeostasis, Fe starvation, and cold stress (Dai et al. 2016). They might act either as activators or repressor of

transcription. Particularly, in maize, different *WRKY* genes confer opposite stress-related phenotypes when over-expressed in Arabidopsis (Cai et al. 2017; Li et al. 2013).

Among down-regulated genes in WST10 compared to NST10, transcription regulators, chromatin-related regulators of transcription, and genes related to DNA synthesis/replication and cell cycle were identified, indicating that the prolonged drought stress caused a block of cell division and expansion in the developing leaves. Concomitantly, the strong enrichment in genes coding for histone proteins, chromatin remodeling factors, and DNA methyltransferases in down-regulated gene category suggests an effective involvement of epigenetic mechanisms in mediating transcriptional response to drought stress.

Interestingly, the amplitude of the stress-misregulated gene set was very different between B73 wild type and *rmr6-1* mutant plants. We proposed that this different behavior is the result of stress-like effect on genome regulation caused by the epiregulator mutation itself, which appears to activate many stress-related genes even in control growth condition (Forestan et al. 2016).

7.2.3 sRNA-Seq Analysis of Drought-Induced Small RNA Variations in Maize Leaves

Stress-induced changes in maize miRNAs and siRNAs accumulation were investigated by sRNA Illumina sequencing of B73 and *rmr6-1* mutant leaves (Lunardon et al. 2016). The majority of genome-aligned sRNAs from B73 leaves were 24 nt long. Conversely, 24-nt sRNAs were almost absent in the *rmr6-1* mutant, where 22-nt sRNAs had a slightly higher accumulation level. No major variation in the sRNA size distribution was observed in the stressed compared to the non-stressed leaves. The merge set of all aligned reads was then used for the de novo annotation of maize sRNA loci that led to the identification of 48 *MIRNA* loci and >250,000 non-*MIRNA* loci with most frequent RNA size between 20 nt and 24 nt. Eighteen out of 48 *MIRNA* loci identified by this stringent analysis were new loci that produced miRNAs either 21 or 22 nt long. As expected, their accumulation was not affected in the *rmr6-1* background.

Differential expression analysis on sRNAs revealed no major global changes in the sRNA profiles of maize leaves following drought stress application. To test more in detail the effects of the applied drought stress treatment on miRNAs, we performed differential expression analysis on the mature miRNAs annotated in miRBase and in the de novo annotation, because changes of even a few microRNAs may be interesting for elucidating the cross-talk between drought stress and plant development in maize. Three miRNAs, miR156, miR2275, and miR398, were up-regulated in B73 leaf by drought stress; conversely, miR166 and miR396 were down-regulated by the stress.

The up-regulation of miR156 was previously observed in maize and also in many other plant species following diverse stress applications. Particularly, in the model *Arabidopsis thaliana,* it has been showed that miR156 negatively regulates *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3* (*AtSPL3*) expression via cleavage of its transcripts (Wu and Poethig 2006). The SPL3 transcription factor promotes the floral transition by activating the expression of *FRUITFULL* (*FUL*), *LEAFY* (*LFY*), and *APETALA1* (*AP1*) (Yamaguchi et al. 2009). SPL3 represses the floral transition, and thus, over-expression of miR156 maintains the plant in the juvenile state for a longer time and delays flowering time. Further investigations are required in maize to assess the role of miR156 in the regulation of flowering time under drought stress conditions. However, in our experiments drought stress treatments were associated with late flowering phenotypes and thus it would be interesting to determine whether miR156 up-regulation might be directly involved in delaying flowering time following drought stress.

miR398 is a highly conserved marker of stress in plants, and many works have revealed that it is involved in responses to diverse abiotic and

biotic stresses (for a review, see Zhu et al. 2011). Although maize target genes of miR398 are not known, in Arabidopsis mRNA targets are copper/zinc superoxide dismutases (CSDs), which are important ROS scavengers (Mittler 2002; Sunkar et al. 2006). In *M. truncatula,* miR398a/b is strongly up-regulated during water deficit and inversely correlated to the expression levels of its targets, among which are CSD1, CSD2, and also COX5b that is involved in electron transport in the mitochondrial respiratory pathway (Trindade et al. 2010). To definitively confirm that miR398 represents a stress marker in maize, it would be necessary to identify and validate its mRNA targets.

The down-regulation of miR166 and miR396 expression in stressed leaves, which are involved in adaxial/abaxial (dorsoventral) leaf polarity determination (Chen 2012; Mecchia et al. 2013; Zhang et al. 2006), indicates that drought stress interferes with leaf development in maize.

Finally, it was observed that the effect of drought stress on miRNA accumulation was different between B73 and *rmr6-1* mutant, which was not expected because miRNAs are transcribed by PolII (Rogers and Chen 2013) and in control conditions no alterations in miRNA expression were detected in *rmr6-1* leaves compared to B73. We concluded that this difference is probably due to the secondary effects of the *rmr6-1* mutation (Lunardon et al. 2016).

Similar to *MIRNA* loci, a small fraction of non-*MIRNA* loci were drought stress responsive in maize leaf, with a bias toward a drought-induced down-regulation of sRNA loci both during the stress and recovery. Only 0.02% of the siRNA loci was differentially expressed, with an equivalent number of 22-nt and 24-nt loci. Interestingly, we found that five *TAS3* loci were down-regulated by drought stress in B73 leaf. *TAS3* is a conserved ta-siRNA family, whose members target the *AUXIN RESPONSE FACTOR 3* and restricts its expression to the abaxial leaf domain (Guan et al. 2017; Peragine et al. 2004). This observation on misregulation of leaf patterning is consistent with the results on miRNAs differential expression in leaf. Furthermore, a recent study highlighted a potential role for maize ta-siRNA, together with histone modifications, in regulation of NAT expression in response to drought stress (Xu et al. 2017), indicating that plant response and adaptation to drought is regulated by complex genetic and epigenetic networks and might require a rapid reprogramming of plant leaf growth.

7.3 Analysis of Chromatin Landscape at Drought Stress-Responsive Genes

Epigenetic components play a major role in plant interaction and adaptation to both non-stressful and stressful environmental conditions by altering the competence of genetic information to be expressed. As in all eukaryotic cells, plant genomic DNA is compacted into chromatin in the nucleus. The nucleosome is the basic structural unit of chromatin and consists of roughly 146 DNA base pairs wrapped around a histone octamer or nucleosome core particle. The nucleosome core particle is made up by two units of each core histone, H2A, H2B, H3, and H4. H3–H4 form a tetramer that organizes the central part of the nucleosomal DNA (about 80 bp), while the two H2A–H2B dimers each bind roughly 40 bp of DNA and constitute the entry/exit points of nucleosomal DNA access. Core histone proteins can be subjected to post-translational modifications and have variant isoforms codified by the genome. In both animals and plants, histone modifications have important role in regulating chromatin dynamics and it has become an accepted dogma that the combination of post-translational modifications of histones, such as acetylation/deacetylation, methylation, phosphorylation, ubiquitination, is indicative of the transcriptional state of a gene sequence in that chromatin context. In plants, among the many identified histone modifications some are commonly associate with transcriptional activation (histone acetylation, trimethylation of histone H3 lysine 4 or 36, H3K4me3, and H3K36me3), while others are associated with gene silencing (histone deacetylation, trimethylation of histone

H3 lysine 27—H3K27me3 or dimethylation of H3 lysine 9, H3K9me2). Along with histone post-translational modifications and deposition of histone variants, DNA methylation occurring at cytosine residues is the best-characterized epigenetic mark in plants and contributes to the transcriptional state of a gene (reviewed in Du et al. 2015; Yuan et al. 2013).

Variation in chromatin structure often accompanies variation in gene expression: according to recent epigenetic models, chromatin remodelers, histone modifiers, and DNA methylating/demethylating activities interact with the mediation of both short and long ncRNAs for regulating adaptation to the environment (Asensi-Fabado et al. 2017; Haak et al. 2017). Furthermore, environmental adaptation requires a fine-tuning between external cues and the timing of plant growth and developmental changes. Because often a certain delay occurs between the environmental trigger and the initiation of a differentiation process, a memory of the trigger can be epigenetically set and reset in each generation (Avramova 2015; Lamke and Baurle 2017). This also means that some transcriptional adjustments for stress adaptation can persist after the end of stress- and trigger-adaptive non-DNA sequence-based mechanisms of transgenerational inheritance.

7.3.1 ChIP-Seq Analysis of Stress-Induced Histone Modification

To investigate the association between gene transcription and chromatin features in response to drought stress, the genome-wide distribution of trimethylation on lysine 4 of histone 3 (H3K4me3) and acetylation on histone 3 lysine 9 (H3K9ac) were analyzed by means of ChIP-Seq. Both chromatin modifications are strongly associated with gene transcriptional activation and are almost exclusively located within the 5′ region of genes, with peaks around the transcription start site (TSS). Chromatin was extracted from the same leaf samples used in RNA-Seq analysis and was immunoprecipitated using specific antibodies that recognized H3K4me3 and H3K9ac prior to libraries preparation for Illumina sequencing.

High-quality reads were mapped to the maize reference genome: uniquely mapped reads were analyzed for each histone mark in each sample and compared with sequenced reads of input control (represented by the whole chromatin which was not immunoprecipitated prior to sequencing). In detail, for each analyzed sample, H3K4me3 and H3K9ac ChIP-Seq read counts were calculated for each gene over the 1 Kb region downstream of the TSS, and multiple 2×2 χ^2 tests were used to identify genes that were significantly enriched in the each histone modification with respect to the input control, as well as to identify genes displaying a consistent gain or loss of histone marks in response to the applied drought stress (Forestan et al., *unpublished*).

Overall, about 27,000 annotated maize genes were evaluated for difference in H3K4me3 compared to the input control, and 1,044 of these genes showed a significant increase of H3K4me3 within the 1 kb region downstream the TSS in WST10 compared to NST10, while 958 genes were associated with a significant loss of this mark. For H3K9ac, the number of genes associated with a significant gain (249) and loss (891) in WST10 vs NST10 was lower than H3K4me3, mainly because of the overall lower number of maize genes significantly marked by acetylation (about 18,000). A high positive correlation between drought-induced significant changes to histone marks and transcriptional variation was observed for both histone modifications (Forestan et al., *unpublished*).

Representative genes displaying significant changes in H3K4me3 or H3K9ac directly correlated with gene transcription variation are shown in Fig. 7.2. HVA22 is an abscisic acid (ABA)/stress-induced protein first isolated from barley (*Hordeum vulgare*) aleurone cells; it is involved in vesicular trafficking, and it ensures tolerance to low water availability during seed germination (Brands and Ho 2002; Shen et al. 2001). Its transcriptional up-regulation was used as a drought stress marker during the first pilot

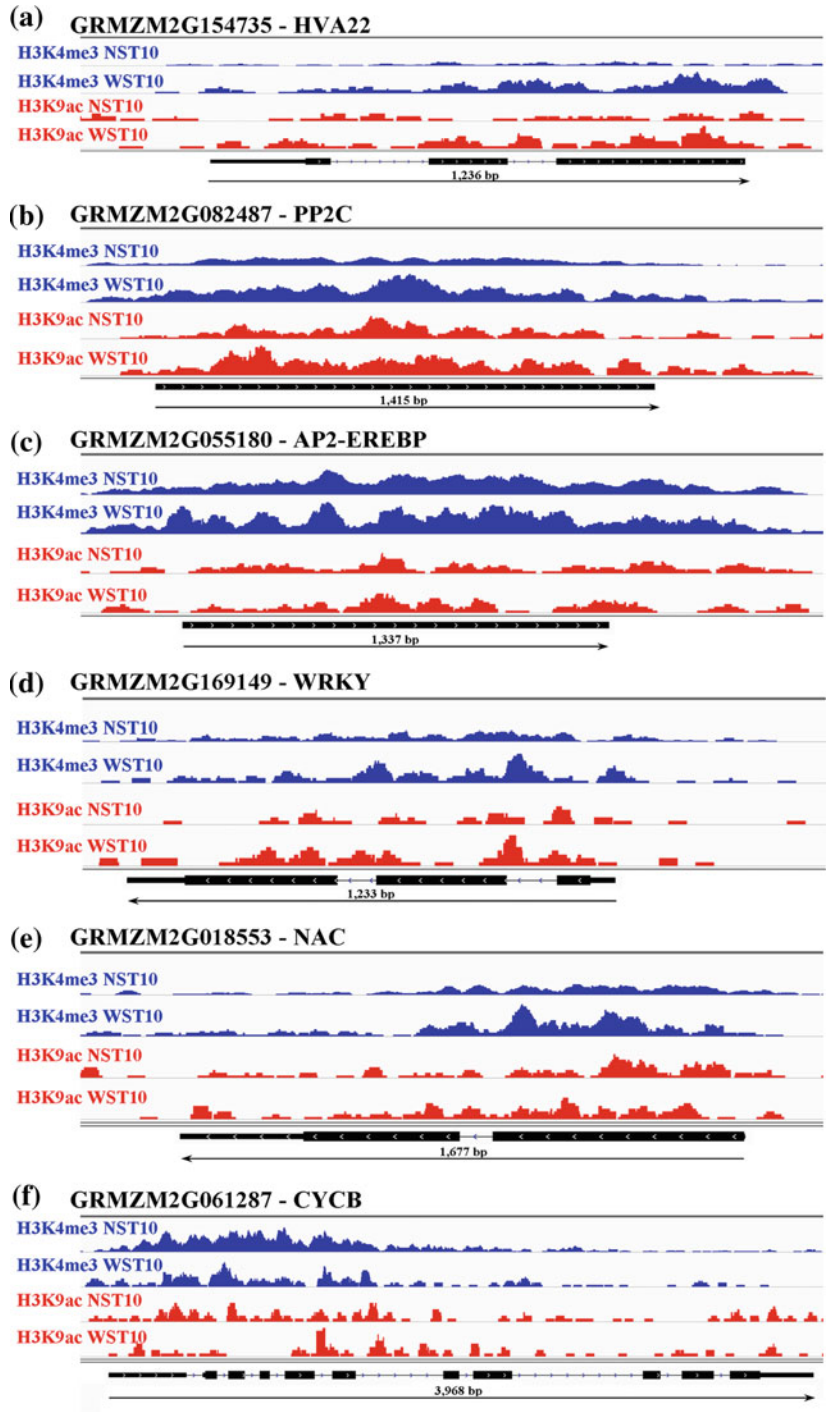

Fig. 7.2 H3K4me3 and H3K9ac histone modifications for selected drought-responsive genes. Integrative Genomics Viewer (IGV; http://software.broadinstitute.org/software/igv/) snapshots reporting mapped reads (in blue for H3K4me3 and red for H3K9ac) and locus annotation (black boxes) for six maize genes up- (**a** to **e**) or down-regulated (**f**) following prolonged drought stress (WST10) compared to control samples (NST10). Coverage of ChIP-Seq reads (normalized to the total number of mapped reads) shows significant stress-induced gain in both H3K4me3 and H3K9ac at the 5′-end regions of stress up-regulated *HVA22* (**a**), *PP2C* (**b**), and *WRKY* (**d**) genes, while only the H3K4me3 histone mark increased in *AP2-EREBP* (**c**) and *NAC* (**e**) coding loci. Conversely, a loss of H3K4me3 was highlighted following stress application in the 5′-end region of the *CYCB* coding gene (**f**)

experiment and has been confirmed also by RNA-Seq analysis (log2 FC WST10/NST10 = 3.95). ChIP-Seq analysis revealed that its strong transcriptional activation is linked to a significant increase in both H3K4me3 and H3K9ac histone marks in the 5′-end of the gene region (Fig. 7.2a).

Similarly, both analyzed histone modifications were strongly enriched at several protein phosphatases 2C (PP2C) coding genes whose transcription was significantly induced by the drought stress (ChIP-Seq reads for GRMZM2G082487, which showed a log2 FC of 2.143 in WST10 versus NST10, are reported in Fig. 7.2b). PP2Cs were shown to function as co-receptors for the phytohormone abscisic acid, representing therefore central components in ABA-dependent osmotic stress signal transduction pathways (Moorhead et al. 2007).

Significant gain in H3K4me3 levels were found also at 5′-end region of loci coding for transcription factors of the previously described stress-responsive TF families, such as AP2-EREBP, WRKY, and NAC TF members (Fig. 7.2c–e), which were transcriptionally activated by the prolonged drought stress. Conversely, drought stress was shown to induce the down-regulation of genes regulating cell cycle and cell division (Fig. 7.1b) and a significant loss in H3K4me3 histone mark was correlated with their transcriptional silencing, such as the case of the Cyclin B2 locus reported in Fig. 7.2f (log2 FC WST10/NST10 = −2.36).

7.4 Conclusions

Drought represents the most serious abiotic stress limiting crop production worldwide; therefore, the development of drought-tolerant crops is a major goal for future sustainable agriculture. To this end, a complete understanding of plants physiological, biochemical, and gene regulatory networks involved in drought perception, response, and adaptation is essential. In this framework, tremendous advances have been made over the past decade in shedding light on the complex mechanisms of drought stress responses in crops, through the use of various omics approaches (including genomics, transcriptomics, proteomics, metabolomics, and epigenomics).

Application and integration of omics assays highlighted the existence of combinatorial gene regulation, which includes different types of ncRNAs and chromatin/histone modifications that underlie maize drought response and tolerance. In particular, integration of RNA-Seq and ChIP-Seq data, unveiled a direct correlation between stress-induced genes transcriptional variation and H3K4me3 or H3K9ac histone modification levels, allowed the identification of a robust list of epigenetic targets that affect different stress-responsive, developmental, and metabolic pathways. Further studies, primarily aimed at the dissection of genotype × environment interactions in complex field experiments, are necessary to deeply characterize these genes and pathways, which may be manipulated to improve crop drought tolerance and productivity in water-limited conditions.

Additional studies on transcriptional and epigenetic responses during the stress-recovery period will investigate whether a stress-memory formation occurs or resetting mechanisms are initiated. In particular, the recovery, the period after the stress that is needed to reach a new homeostasis, represents a critical point for studying whether resetting mechanisms are initiated after the stress or if memory formation occurs. Histone H3K4 methylation was indeed frequently correlated with different types of somatic stress memory in Arabidopsis, but insights into the molecular conservation of stress memory in crop species remain scarce. Understanding both the molecular bases and the crop-specific targets of stress memory will ultimately enable the improvement of stress tolerance in crops.

Acknowledgements Thanks to Vincenzo Rossi and Giulio Pavesi for their precious support on ChIP-Seq. This work was conducted within the grants from the European Commission (FP7 Project KBBE 2009 226477 —"AENEAS": Acquired Environmental Epigenetics

Advances: from Arabidopsis to maize) and Italian MIUR-CNR Flagship project EPIGEN.

References

Abdul Jaleel C, Manivannan P, Wahid A, Farooq M, Jasim Al-Juburi H, Somasundaram R, Panneerselvam R (2009) Drought stress in plants: a review on morphological characteristics and pigments composition

Agarwal PK, Gupta K, Lopato S, Agarwal P (2017) Dehydration responsive element binding transcription factors and their applications for the engineering of stress tolerance. J Exp Bot 68:2135–2148

Asensi-Fabado MA, Amtmann A, Perrella G (2017) Plant responses to abiotic stress: the chromatin context of transcriptional regulation. Biochim Biophys Acta 1860:106–122

Augustine SM, Ashwin Narayan J, Syamaladevi DP, Appunu C, Chakravarthi M, Ravichandran V, Tuteja N, Subramonian N (2015) Overexpression of EaDREB2 and pyramiding of EaDREB2 with the pea DNA helicase gene (PDH45) enhance drought and salinity tolerance in sugarcane (Saccharum spp. hybrid). Plant Cell Rep 34:247–263

Avramova Z (2015) Transcriptional 'memory' of a stress: transient chromatin and memory (epigenetic) marks at stress-response genes. Plant J 83:149–159

Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64:137–159

Banziger M, Araus J (2007) Recent advances in breeding maize for drought and salinity stress tolerance. In: Jenks MA, Hasegawa PM, Jain SM (eds) Advances in molecular breeding toward drought and salt tolerant crops. Springer, Netherlands, Dordrecht, pp 587–601

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, Sanmiguel PJ, Bennetzen JL (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet 5:e1000732

Blum A (2014) Genomics for drought resistance—getting down to earth

Brands A, Ho TH (2002) Function of a plant stress-induced gene, HVA22. Synthetic enhancement screen with its yeast homolog reveals its role in vesicular traffic. Plant Physiol 130:1121–1131

Cai R, Dai W, Zhang C, Wang Y, Wu M, Zhao Y, Ma Q, Xiang Y, Cheng B (2017) The maize WRKY transcription factor ZmWRKY17 negatively regulates salt stress tolerance in transgenic Arabidopsis plants. Planta 246:1215–1231

Chen WJ, Zhu T (2004) Networks of transcription factors with roles in environmental stress response. Trends Plant Sci 9:591–596

Chen X (2012) Small RNAs in development—insights from plants. Curr Opin Genet Dev 22:361–367

Chung PJ, Jung H, Jeong DH, Ha SH, Choi YD, Kim JK (2016) Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice. BMC Genomics 17:563-016-2997-3

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17:13-016-0881-8

Cook BI, Anchukaitis KJ, Touchan R, Meko DM, Cook ER (2016) Spatiotemporal drought variability in the Mediterranean over the last 900 years 121:2060–2074

Cui M, Zhang W, Zhang Q, Xu Z, Zhu Z, Duan F, Wu R (2011) Induced over-expression of the transcription factor OsDREB2A improves drought tolerance in rice. Plant Physiol Biochem 49:1384–1391

Dai X, Wang Y, Zhang WH (2016) OsWRKY74, a WRKY transcription factor, modulates tolerance to phosphate starvation in rice. J Exp Bot 67:947–960

Dey S, Corina Vlot A (2015) Ethylene responsive factors in the orchestration of stress responses in monocotyledonous plants. Front Plant Sci 6:640

Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, Zhang T, Qi Y, Gerstein MB, Guo Y, Lu ZJ (2014) Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. Plant J 80:848–861

Ding ZJ, Yan JY, Xu XY, Yu DQ, Li GX, Zhang SQ, Zheng SJ (2014) Transcription factor WRKY46 regulates osmotic stress responses and stomatal movement independently in Arabidopsis. Plant J 79:13–27

Du J, Johnson LM, Jacobsen SE, Patel DJ (2015) DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol 16:519–532

Erhard KF Jr, Talbot JE, Deans NC, McClish AE, Hollick JB (2015) Nascent transcription affected by RNA polymerase IV in Zea mays. Genetics 199:1107–1125

Erhard KF, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB (2009) RNA polymErase IV functions in paramutation in Zea mays. Science 323:1201–1205

Forestan C, Aiese Cigliano R, Farinati S, Lunardon A, Sanseverino W, Varotto S (2016) Stress-induced and epigenetic-mediated maize transcriptome regulation study by means of transcriptome reannotation and differential expression analysis. Sci Rep 6:30446

Guan C, Wu B, Yu T, Wang Q, Krogan NT, Liu X, Jiao Y (2017) Spatial auxin signaling controls leaf flattening in Arabidopsis. Curr Biol 27(2940–2950):e4

Haak DC, Fukao T, Grene R, Hua Z, Ivanov R, Perrella G, Li S (2017) Multilevel regulation of abiotic stress responses in plants. Front Plant Sci 8:1564

Harrison MT, Tardieu F, Dong Z, Messina CD, Hammer GL (2014) Characterizing drought stress and trait influence on maize yield under current and future conditions. Glob Chang Biol 20:867–878

Hong Y, Zhang H, Huang L, Li D, Song F (2016) Overexpression of a stress-responsive NAC

transcription factor gene ONAC022 improves drought and salt tolerance in rice. Front Plant Sci 7:4

Iwaki T, Guo L, Ryals JA, Yasuda S, Shimazaki T, Kikuchi A, Watanabe KN, Kasuga M, Yamaguchi-Shinozaki K, Ogawa T, Ohta D (2013) Metabolic profiling of transgenic potato tubers expressing Arabidopsis dehydration response element-binding protein 1A (DREB1A). J Agric Food Chem 61:893–900

Jia J, Fu J, Zheng J, Zhou X, Huai J, Wang J, Wang M, Zhang Y, Chen X, Zhang J, Zhao J, Su Z, Lv Y, Wang G (2006) Annotation and expression profile analysis of 2073 full-length cDNAs from stress-induced maize (Zea mays L.) seedlings. Plant J 48:710–727

Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, Pareek A, Singla-Pareek SL (2016) Transcription factors and plants response to drought stress: current understanding and future directions. Front Plant Sci 7:1029

Kumar V, Khare T, Shriram V, Wani SH (2018) Plant small RNAs: the essential epigenetic regulators of gene expression for salt-stress responses and tolerance. Plant Cell Rep 37:61–75

Lamke J, Baurle I (2017) Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biol 18:124-017-1263-6

Laraus J (2004) The problems of sustainable water use in the Mediterranean and research requirements for agriculture. Ann Appl Biol 144:259–272

Lembke CG, Nishiyama MY Jr, Sato PM, de Andrade RF, Souza GM (2012) Identification of sense and antisense transcripts regulated by drought in sugarcane. Plant Mol Biol 79:461–477

Li H, Gao Y, Xu H, Dai Y, Deng D, Chen J (2013) ZmWRKY33, a WRKY maize transcription factor conferring enhanced salt stress tolerances in Arabidopsis

Liu J, Wang H, Chua NH (2015) Long noncoding RNA transcriptome of plants. Plant Biotechnol J 13:319–328

Liu S, Wang X, Wang H, Xin H, Yang X, Yan J, Li J, Tran LS, Shinozaki K, Yamaguchi-Shinozaki K, Qin F (2013) Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of Zea mays L. PLoS Genet 9:e1003790

Lobell DB, Roberts MJ, Schlenker W, Braun N, Little BB, Rejesus RM, Hammer GL (2014) Greater sensitivity to drought accompanies maize yield increase in the U. S Midwest. Science 344:516–519

Lu HF, Dong HT, Sun CB, Qing DJ, Li N, Wu ZK, Wang ZQ, Li YZ (2011) The panorama of physiological responses and gene expression of whole plant of maize inbred line YQ7-96 at the three-leaf stage under water deficit and re-watering. Theor Appl Genet 123:943–958

Lunardon A, Forestan C, Farinati S, Axtell M, Varotto S (2016) Genome-wide characterization of maize small RNA loci and their regulation in the required to maintain repression6-1 (rmr6-1) mutant and long-term abiotic stresses. Plant Physiol 170:1535–1548

Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet 11:e1004915

Mao H, Wang H, Liu S, Li Z, Yang X, Yan J, Li J, Tran LS, Qin F (2015) A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nat Commun 6:8326

Matsui A, Nguyen AH, Nakaminami K, Seki M (2013) Arabidopsis non-coding RNA regulation in abiotic stress responses. Int J Mol Sci 14:22642–22654

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15:394–408

Mecchia MA, Debernardi JM, Rodriguez RE, Schommer C, Palatnik JF (2013) MicroRNA miR396 and RDR6 synergistically regulate leaf development. Mech Dev 130:2–13

Miao Z, Han Z, Zhang T, Chen S, Ma C (2017) A systems approach to a spatio-temporal understanding of the drought stress response in maize. Sci Rep 7:6590-017-06929-y

Mittler R (2002) Oxidative stress, antioxidants and stress tolerance. Trends Plant Sci 7:405–410

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. Biochim Biophys Acta 1819:86–96

Moorhead GB, Trinkle-Mulcahy L, Ulke-Lemee A (2007) Emerging roles of nuclear protein phosphatases. Nat Rev Mol Cell Biol 8:234–244

Morari F, Meggio F, Lunardon A, Scudiero E, Forestan C, Farinati S, Varotto S (2015) Time course of biochemical, physiological, and molecular responses to field-mimicked conditions of drought, salinity, and recovery in two maize lines. Front Plant Sci 6:314

Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front Plant Sci 5:170

Okay S, Derelli E, Unver T (2014) Transcriptome-wide identification of bread wheat WRKY transcription factors in response to drought stress. Mol Genet Genomics 289:765–781

Opitz N, Paschold A, Marcon C, Malik WA, Lanz C, Piepho HP, Hochholdinger F (2014) Transcriptomic complexity in young maize primary roots in response to low water potentials. BMC Genomics 15:741-2164-15-741

Paul S, Gayen D, Datta SK, Datta K (2015) Dissecting root proteome of transgenic rice cultivars unravels metabolic alterations and accumulation of novel stress responsive proteins under drought stress. Plant Sci 234:133–143

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are

required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev 18:2368–2379

Phukan UJ, Jeena GS, Shukla RK (2016) WRKY transcription factors: molecular regulation and stress responses in plants. Front Plant Sci 7:760

Qin T, Zhao H, Cui P, Albesher N, Xiong L (2017) A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. Plant Physiol 175:1321–1336

Qiu Y, Yu D (2009) Over-expression of the stress-induced OsWRKY45 enhances disease resistance and drought tolerance in Arabidopsis. Environ Exp Bot 65:35–47

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biol Chem 379:633–646

Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell 25:2383–2399

Seager R, Tzanova A, Nakamura J (2009) Drought in the Southeastern United States: causes, variability over the last millennium, and the potential for future hydroclimate change. J Climate 22:5021–5045

Shan X, Li Y, Jiang Y, Jiang Z, Hao W, Yuan Y (2013) Transcriptome profile analysis of maize seedlings in response to high-salinity, drought and cold stresses by deep sequencing. Plant Mol Biol Rep 31:1485–1491

Shavrukov Y, Baho M, Lopato S, Langridge P (2016) The TaDREB3 transgene transferred by conventional crossings to different genetic backgrounds of bread wheat improves drought tolerance. Plant Biotechnol J 14:313–322

Shen Q, Chen CN, Brands A, Pan SM, Ho TH (2001) The stress—and abscisic acid-induced barley gene HVA22: developmental regulation and homologues in diverse organisms. Plant Mol Biol 45:327–340

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58:221–227

Singh B, Singh B, Kumar V, Pankaj Kumar S, Jayaswal PK, Shefali M, Singh N (2015) Haplotype diversity and association analysis of SNAC1 gene in wild rice germplasm

Sunkar R, Kapoor A, Zhu JK (2006) Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance. Plant Cell 18:2051–2065

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578

Trindade I, Capitao C, Dalmay T, Fevereiro MP, Santos DM (2010) miR398 and miR408 are up-regulated in response to water deficit in Medicago truncatula. Planta 231:705–716

Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, maize. Plant Cell Environ 32:1211–1229

Wang D, Qu Z, Yang L, Zhang Q, Liu ZH, Do T, Adelson DL, Wang ZY, Searle I, Zhu JK (2017a) Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. Plant J 90:133–146

Wang H, Wang H, Shao H, Tang X (2016) Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. Front Plant Sci 7:67

Wang J, Meng X, Dobrovolskaya OB, Orlov YL, Chen M (2017b) Non-coding RNAs and their roles in stress response in plants. Genomics Proteomics Bioinform. 15:301–312

Wu G, Poethig RS (2006) Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. Development 133:3539–3547

Wu S, Ning F, Zhang Q, Wu X, Wang W (2017) Enhancing omics research of crop responses to drought under field conditions. Front Plant Sci 8:174

Wu X, Shiroto Y, Kishitani S, Ito Y, Toriyama K (2009) Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing OsWRKY11 under the control of HSP101 promoter. Plant Cell Rep 28:21–30

Xu J, Wang Q, Freeling M, Zhang X, Xu Y, Mao Y, Tang X, Wu F, Lan H, Cao M, Rong T, Lisch D, Lu Y (2017) Natural antisense transcripts are significantly involved in regulation of drought stress in maize. Nucleic Acids Res 45:5126–5141

Yamaguchi A, Wu MF, Yang L, Wu G, Poethig RS, Wagner D (2009) The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1. Dev Cell 17:268–278

You J, Zong W, Hu H, Li X, Xiao J, Xiong L (2014) A STRESS-RESPONSIVE NAC1-regulated protein phosphatase gene rice protein phosphatase18 modulates drought and oxidative stress tolerance through abscisic acid-independent reactive oxygen species scavenging in rice. Plant Physiol 166:2100–2114

Yu S, Ligang C, Liping Z, Diqiu Y (2010) Overexpression of OsWRKY72 gene interferes in the abscisic acid signal and auxin transport pathway of Arabidopsis. J Biosci 35:459–471

Yuan L, Liu X, Luo M, Yang S, Wu K (2013) Involvement of histone modifications in plant abiotic stress responses. J Integr Plant Biol 55:892–901

Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. Plant J 46:243–259

Zhang W, Han Z, Guo Q, Liu Y, Zheng Y, Wu F, Jin W (2014) Identification of maize long non-coding RNAs responsive to drought stress. PLoS ONE 9:e98958

Zhang YC, Chen YQ (2013) Long noncoding RNAs: new regulators in plant development. Biochem Biophys Res Commun 436:111–114

Zhao L, Wang P, Yan S, Gao F, Li H, Hou H, Zhang Q, Tan J, Li L (2014) Promoter-associated histone acetylation is involved in the osmotic stress-induced transcriptional regulation of the maize ZmDREB2A gene. Physiol Plant 151:459–467

Zhu C, Ding Y, Liu H (2011) MiR398 and plant stress responses. Physiol Plant 143:1–9

8 Maize Small RNAs as Seeds of Change and Stability in Gene Expression and Genome Stability

Reza Hammond, Chong Teng and Blake C. Meyers

Abstract

Studies of maize small RNAs (sRNAs) are providing insights and novel discoveries in both RNA biology and plant evolution. With improvements to the genome and the development of a broad array of resources supporting maize research, maize sRNAs have become better understood in recent years. Here, we review the major classes of maize sRNAs, including heterochromatic small interfering RNAs (hc-siRNAs), phased secondary small interfering RNAs (phasiRNAs), and microRNAs (miRNAs). We examine their biogenesis, roles in paramutation, and functions, including both what is known and hypothesized. Finally, we describe the resources for their study, including bioinformatics tools that will contribute in future studies of maize sRNAs to the elucidation of their biogenesis, function, and evolution.

8.1 Overview of Maize Small RNAs

As their name suggests, endogenous small RNAs (sRNA) in plants are short in length, 21–24 nucleotides, and thus are unable to encode proteins. Instead, small RNAs guide regulatory silencing processes, mostly on the basis of sequence complementarity, by guiding ARGONAUTE (AGO) proteins to the targets. Small RNAs function at either the transcriptional (DNA) or posttranscriptional (RNA) level, and they are found widely in eukaryotes including yeast, animals, and plants (Borges and Martienssen 2015). Small RNAs are typically processed from double-stranded or hairpin RNA structures by RNase III enzymes (Dicer-like proteins, or DCLs) from transcripts made by different RNA polymerases. After processing, the sRNAs are loaded into AGOs to perform their silencing functions. Although sRNAs are indeed small in size, they are complex in their biogenesis and functions, especially in plants (Fig. 8.1). In *Arabidopsis*, for example, there are three

Reza Hammond and Chong Teng are equally contributed.

R. Hammond
Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19714, USA

R. Hammond
Delaware Biotechnology Institute, University of Delaware, Newark, DE 19714, USA

C. Teng · B. C. Meyers
Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

B. C. Meyers (✉)
52 Agriculture Lab, Division of Plant Sciences, University of Missouri – Columbia, Columbia, MO 65211, USA
e-mail: bmeyers@danforthcenter.org

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_8

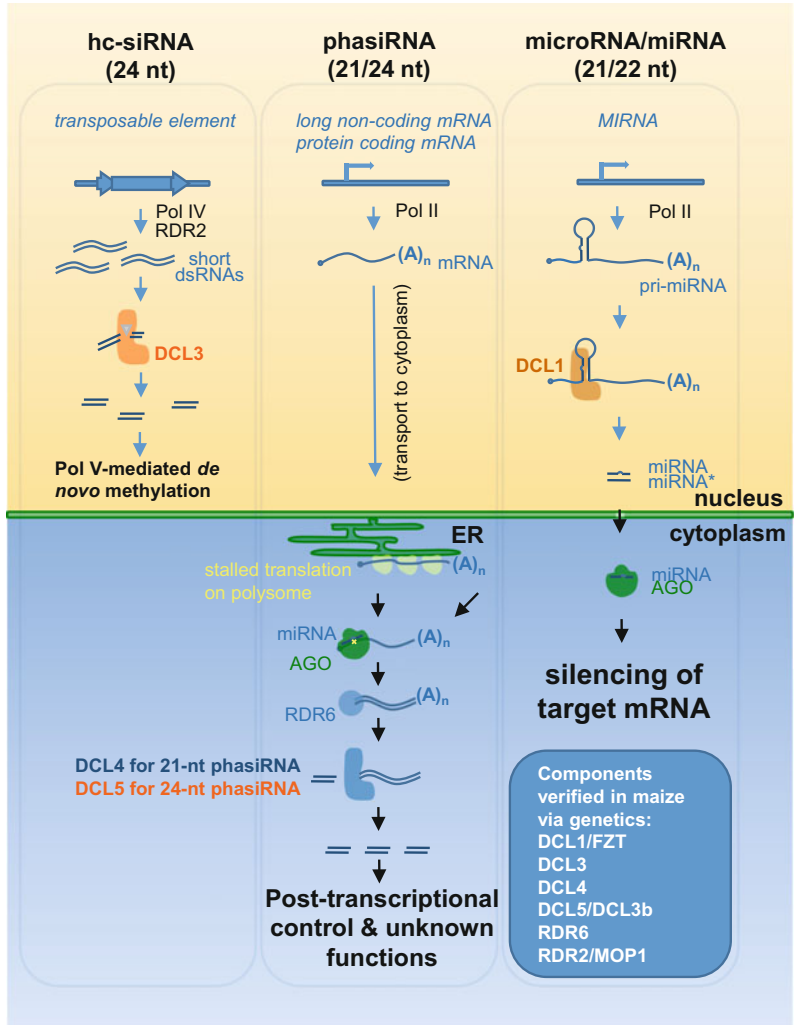

Fig. 8.1 Overview of small RNA biogenesis pathways in maize. There are three well-characterized categories of small RNAs: microRNAs (or "miRNAs"), typically 21–22 nucleotides (nt) long; heterochromatic small interfering RNAs (hc-siRNAs), 23–24 nt in length; and phased secondary small interfering RNA (phasiRNAs), mainly 21 or 24 nt in length. hc-siRNAs are diced in the nucleus by DCL3 from short dsRNA precursors, products of Pol IV and RDR2. Then, the hc-siRNAs will direct Pol V-dependent de novo methylation at their targeting loci based on sequence complementary. phasiRNAs are generated on membrane-bound polysomes in the cytoplasm from mRNA precursors including long noncoding RNAs or protein-coding mRNAs produced by Pol II. miRNA–AGO complexes initiate a first cut on the phasiRNA precursors, triggering dsRNA conversion by RDR6. Then, DCL4 and (monocot-specific) DCL5 dice the dsRNA into 21- or 24-nt phasiRNAs, respectively. miRNAs are generated as noncoding RNA products of Pol II. The Pol II-derived precursor miRNAs are first folded into imperfect matched hairpin structures, then diced by DCL1 with help from other necessary cofactors (HYL1, SGS3, etc.) yielding miRNA-miRNA* duplexes. Then, the functional miRNAs are loaded into cognate AGO proteins (predominantly AGO1) and translocated into cytosol to perform their function on mRNAs. Several key components mentioned here and in the main text have been validated by years of study in maize mutants, as indicated in the blue box on the bottom right

well-characterized categories of small RNAs: microRNA (miRNAs), typically 21 or 22 nucleotides (nt) in length; heterochromatic small interfering RNAs (hc-siRNAs), ~24 nt in length; and *trans*-acting small interfering RNAs (tasiRNAs), mainly 21 nt in length.

In *Arabidopsis*, each of these sRNA classes is produced and functions via complex and distinct interactions of the existing four Dicer-like proteins (DCLs), ten AGOs, six RNA-dependent RNA polymerases (RDRs), and three RNA polymerase complexes encoded in the *Arabidopsis* genome.

The mechanisms of sRNA biogenesis and silencing in *Arabidopsis* have been thoroughly studied over the past two decades. Due to the conservation of many of these pathways across plants, much (but not all) of what has been uncovered in *Arabidopsis* also pertains to sRNAs found in other plant species. While the small size of the *Arabidopsis* genome is sometimes an advantage in sRNA studies, it is not without limitations. In many other plant species, including maize, each of the three major classes of sRNAs is more diverse, and some families of enzymes involved in the production of sRNAs have expanded roles. As one example, a consequence of the larger, more repetitive genomes of plants such as maize is that they produce diverse populations of hc-siRNAs from large transposable element populations. Small RNAs in maize also have more diverse functions than in *Arabidopsis*, as they have been implicated in phenomena such as paramutation that are well described in maize (Hollick 2016). And as studies of sRNAs in maize and other monocots have intensified, additional sRNA-generating loci have been characterized, yielding large sets of long noncoding mRNAs that generate phased secondary small interfering RNA (phasiRNAs) (Johnson et al. 2009; Zhai et al. 2015). This led to the identification of at least one additional and seemingly monocot-specific protein, a fifth Dicer (DCL5) with a unique role, demonstrating that there is diversification of sRNA pathways even within the angiosperms. The identification of diverse phenotypes in maize mutants disrupted in sRNA production or utilization has led to ongoing interest in complete characterization of these pathways and their roles. In this review, we summarize the current understanding of what makes maize both similar and distinct in its small RNA pathways and the future directions of small RNA-focused studies in maize.

8.2 Heterochromatic, Small Interfering RNAs

8.2.1 Overview of Heterochromatic siRNAs

We start with the most diverse and often most abundant class of sRNAs, heterochromatic siRNAs (hc-siRNAs), which were initially characterized from experiments performed mostly in *Arabidopsis*. hc-siRNAs are ~24 nucleotides in length and are typically derived in plant genomes from transposable elements (TEs). hc-siRNAs direct de novo RNA-directed DNA methylation (RdDM), thus influencing chromatin homeostasis and reinforcing heterochromatin formation and maintenance, and, to a lesser extent, regulating gene activity (Law and Jacobsen 2010; Matzke and Mosher 2014). The canonical RdDM pathway requires plant-specific RNA polymerase complexes, Pol IV and Pol V, both of which evolved from Pol II. The many components of the hc-siRNA biogenesis and RdDM pathway were recently reviewed in detail (Matzke and Mosher 2014), so we will keep the description to a minimum here.

The RdDM pathway follows three major steps: (i) Pol IV-dependent siRNA biogenesis, (ii) Pol V-mediated de novo DNA methylation, and (iii) chromatin alterations facilitated by Pol V via CG, CHG, and CHH DNA methylation marks (Fig. 8.2). Briefly, Pol IV is recruited to transposon-like regions marked by H3K9me where it transcribes hc-siRNA precursors; this recruitment occurs with the help of other protein components like SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1) and CLASSY1 (CLSY1). Next, the precursors are converted into double-stranded RNA (dsRNA) by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). These dsRNA molecules are then cleaved by DICER-LIKE 3 (DCL3). The 24-nt siRNA products are then loaded into ARGONAUTE 4 (AGO4) to direct their function. Upon binding the mature 24-nt hc-siRNA products, AGO4 is recruited by Pol V via proteins that include SUVH2/9, SPT5L/KTF1, IDN2, and the DRD1–DMS3–RDM1 (DDR) complex to target

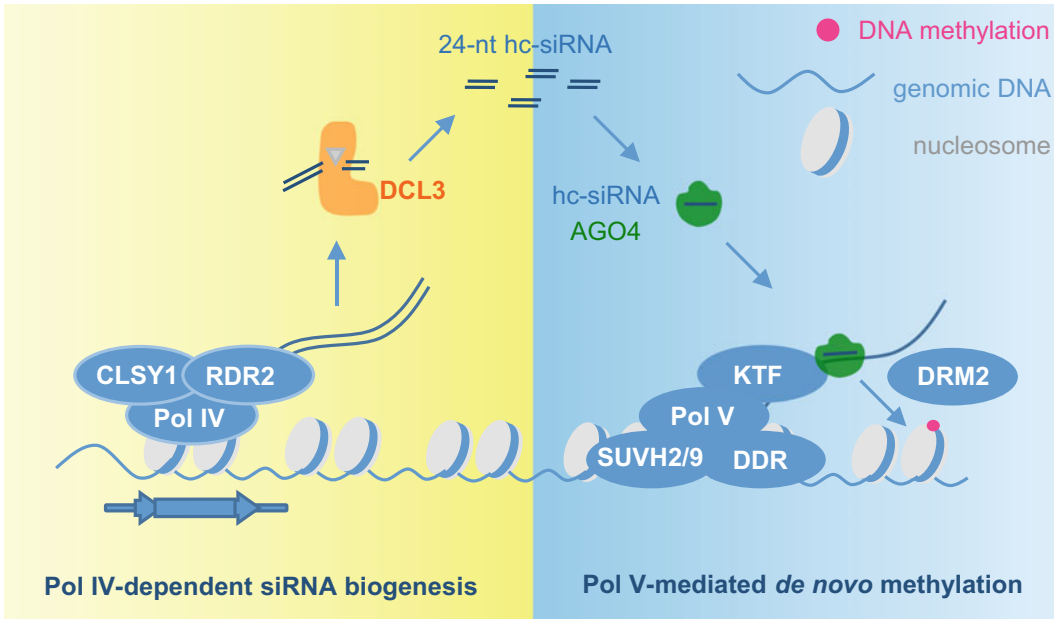

Fig. 8.2 Overview of hc-siRNA and the RdDM pathway in maize. The primary role of hc-siRNAs is in the RdDM pathway, as outlined here, modified from Matzke and Mosher (2014). hc-siRNA biogenesis begins with the transcription of a transposable element repeat by Pol IV and other necessary factors. This product is then converted to a dsRNA by RDR2. Then, the dsRNA precursor is cleaved by DCL3 into a 24-nt hc-siRNA. Some hc-siRNAs are bound by AGO4, the major effector in the canonical RdDM pathway. Finally, the siRNA-AGO4 complex binds to Pol V transcripts via sequence complementary to direct de novo DNA methylation by DRM2. Analyses of mutants and protein–protein interactions in maize has validated effectors in RdDM, including the following: *RMR1*, *RMR1-Like*, *MOP2*/*RMR7*, *MOP3*/*RMR6*, *RDR2*/*MOP1*, *DCL3*/*RMR5*, *AGO4*/*AGO121*

transcripts by sequence homology. When AGO4 finds the target, DOMAINS REARRANGED METHYLASE 2 (DRM2) is recruited by AGO4 and RDM1 to initiate de novo DNA methylation. Finally, the RdDM machinery further influences chromatin as a result of interactions between Pol V scaffold RNAs and the SWI/SNF chromatin remodeling complex, plus an interaction between methylated DNA and histone markers that strengthens gene silencing in a self-reinforcing loop, utilizing the histone-modifying enzymes SUVH4/5/6, UBP26, JMJ14, LDL1/2, and HDA6. Other than this "canonical RdDM," there are a number of effectors of importance that participate in posttranscriptional gene silencing (PTGS) to also effect RdDM; these include RDR6, Pol II, and NERD, which have roles in cases deemed "non-canonical RdDM," reviewed in detail in a subsequent section on phasiRNAs (see Sect. 4). Although only a small proportion of 24 nt hc-siRNAs have been studied in detail (Wang and Axtell 2017), the following section describes the observation that the absence of hc-siRNAs results in a variety of phenotypic defects, arguably accentuated in maize relative to other plants.

8.2.2 Heterochromatic siRNAs in Maize

As the most abundant small RNA species in plants, 24-nt hc-siRNAs have the potential to contribute greatly to phenotypic plasticity. Though not many mutants in RdDM have been isolated in maize, unlike *Arabidopsis*, loss-of-function mutations in maize *RDR2* (originally named *MEDIATOR OF PARAMUTATION 1* (*MOP1*) (Dorweiler et al.

2000), *NRPD1* (*NUCLEAR RNA-DEPENDENT RNA POLYMERASE IV SUBUNIT 1*, originally named *REQUIRED TO MAINTAIN REPRESSION 6, RMR6*) (Hollick et al. 2005), and *DRD1* (aka *RMR1*) (Hale et al. 2007) all substantially impact the production of 24-nt hc-siRNAs, as in *Arabidopsis*. Studies of these maize mutants have revealed roles for hc-siRNAs in paramutation, epiallele regulation, and many other traits, the loss of which is visually apparent (Hollick 2016). Therefore, the developmental phenotypic defects of a loss of single components of the hc-siRNA or RdDM pathway are weaker in *Arabidopsis* than in maize, suggesting that hc-siRNAs have more functional significance in maize. Perhaps this is due to the vastly increased repertoire of repetitive elements in maize, comprising the bulk of its large genome. Alternatively, there might be additional required functions for hc-siRNA-dependent pathways in maize that are less important in *Arabidopsis*. Given the variety of phenotypes more evident in maize due to the loss of hc-siRNAs, more questions about 24-nt hc-siRNAs in maize remain to be addressed.

8.3 Paramutation

8.3.1 Overview of Paramutation

Paramutation is the non-Mendelian segregation of alleles due to an epigenetic change induced by one allele in another allele of the same gene. In such an interaction, the paramutagenic allele produces a heritable change in the interaction of the homologous allele, referred to as the "paramutable" allele (Arteaga-Vazquez et al. 2010; Hollick 2016). In this scenario, crossing a plant carrying a paramutagenic allele with a plant carrying a paramutable allele will result in progeny with suppressed expression of the paramutable allele. Outcrossing these hybrids to plants carrying the paramutable allele results in most, if not all, progeny and successive generations displaying the phenotype of the paramutagenic allele despite not actually carrying this paramutagenic allele. The phenomenon of paramutation was first described by Alexander Brink in 1956 when he identified the maize *red1* gene, which defies the rules of Mendelian inheritance (Arteaga-Vazquez et al. 2010; Brink 1956). Perhaps one of the more curious features of paramutation is that the paramutable allele becomes paramutagenic upon exposure to another paramutagenic allele in *trans* (Hollick 2016).

Paramutation is best understood in maize as the result of studies of four distinct loci: *booster* (*b*), *red* (*r*), *purple plant1* (*pl1*), and *pericarp color1* (*p1*). Each of these genes encodes a transcription factor that regulates the production of anthocyanin pigments. These loci are studied because they are both susceptible to paramutation and they regulate an easy-to-see phenotype—the coloration of various plant tissues. The combined effects of the *b* and *pl1* loci direct plant color, the combined effects of *pl1* and *r* direct anther color, and the *p1* locus directs pericarp color. Despite the similarities in phenotypic consequences, it appears that different mechanisms control paramutation at each of these loci (Hollick 2016).

As a result of mutant screens involving these loci, numerous genes required for paramutation have been identified from multiple labs, identified via forward-genetic screens. Mutants identified as *mediator of paramutation* (*mop*) have been isolated in the *b1* system (Arteaga-Vazquez et al. 2010), while several other mutants identified as *required to maintain repression* (*rmr*) have been isolated in the *pl1* system (Hollick and Chandler 2001). Thus far, all or nearly all genes characterized as required for paramutation apparently encode proteins associated with the biogenesis or function of siRNAs (Table 8.1), which is why paramutation is an important part of this chapter.

8.3.2 siRNAs and Paramutation

The first maize gene cloned for which the loss of function eliminates paramutation was *mop1*, subsequently referred to as $rdr2_{mop1}$. $rdr2_{mop1}$ is a mutation in the maize ortholog of *Arabidopsis thaliana RDR2*, a critical protein in the biogenesis pathway of 24-nt heterochromatic siRNAs (Alleman et al. 2006). In the pathway for

Table 8.1 Loci implicated in paramutation and the mutants identified to suppress paramutation

Arabidopsis gene name	Maize paramutation mutant	Marker for loss of paramutation	References
RDR2	*mop1*	*b* and *pl1*	Dorweiler et al. (2000)
NRPD1	*mop3* and *rmr6*	*b* and *pl1*	Erhard et al. (2009), Sloan et al. (2014)
NRPD2/E2	*mop2* and *rmr7*	*p1*	Sidorenko et al. (2009), Stonaker et al. (2009)
None	*rmr2*	*p1* and r	Barbour et al. (2012)
SNF2/DRD1	*rmr1*	*p1* and r	Hale et al. (2009)
DCL3	*rmr5*	*b*	Gabriel et al. (2015)

production of heterochromatic siRNAs, RDR2 is responsible for the biogenesis of the second strand of the dsRNA precursor of hc-siRNAs (Xie et al. 2004). The $rdr2_{mop1}$ mutant impairs this process and thus largely or completely eliminates the production of hc-siRNAs (Nobuta et al. 2008).

Following the identification of $rdr2_{mop1}$, several other mutants were identified in maize that are deficient in paramutation, all of which encode various proteins involved in the heterochromatic siRNAs biogenesis and/or silencing pathways. In *Arabidopsis*, *NRPD1* encodes the largest subunit of DNA-dependent RNA polymerase IV (Pol IV) (Luo and Hall 2007), and *NRPD2/E2* (*NUCLEAR RNA-DEPENDENT RNA POLYMERASE IV/V SUBUNIT 2*) encodes the shared second largest subunit of Pol IV and Pol V (Pikaard et al. 2008). Pol IV is responsible for the transcription of the first strand of hc-siRNA precursor molecules, so impairing the maize ortholog of *nrpd1* known as *rmr6/mop3* (the dual name resulting from the identification of the same gene in both screens, *rmr* from the Hollick lab and *mop* from the Chandler lab) prevents the accumulation of 24-nt hc-siRNAs, similar to $rdr2_{mop1}$ (Erhard et al. 2009).

The *NRPD2/E2* gene is important to both Pol IV and V. Pol IV is instrumental in the biogenesis of hc-siRNAs, while Pol V is instrumental in the de novo methylation of hc-siRNA targets, as mentioned above. With a mutation expected to simultaneously impede the function of two key proteins in this pathway, the anticipated effect of a maize *nrpd2/e2* mutation would be a significant of a reduction of hc-siRNAs as the other mutants deficient in paramutation. Indeed, accumulation of hc-siRNAs was diminished in maize *nrpd2/e2* mutants, but interestingly, the phenotypic defects of these mutants are not nearly as severe as $nrpd1_{rmr6}$ mutants. The hypothesized reason for the lack of severe phenotypic defects is due to the presence of three presumed functional copies of the *NRPD2/E2*-like genes in maize. Both groups that have published *nrpd2/e2* mutants contain mutations in the same, and only one, *NRPD2/E2*-like gene (Sidorenko et al. 2009; Stonaker et al. 2009). If these genes were fully redundant, loss-of-function mutants would be needed for all copies to see mutant effects. However, the fact that these phenotypic defects and loss of 24-nt hc-siRNAs are observed suggests that these gene copies are not entirely redundant (Pikaard and Tucker 2009).

8.4 Phased, Secondary, Small Interfering RNAs (phasiRNAs)

8.4.1 Overview of phasiRNAs

Phased, secondary, small interfering RNAs (phasiRNAs) are named for their biogenesis, the process which yields a precise head-to-tail arrangement of sRNAs via Dicer processing, starting from the site of cleavage by a miRNA-bound AGO (Borges and Martienssen 2015; Fei et al. 2013). This pathway is potentially powerful because it represents the

amplification of sRNAs, from one miRNA to many phasiRNAs. The well-studied 21- and 22-nt *trans*-acting small interfering RNAs (tasiRNAs) are a good case study (Fei et al. 2013). tasiRNAs were first identified in *Arabidopsis* as secondary products of AGO1/7-bound miRNAs targeting a set of long, noncoding RNAs (lncRNAs) that turned out to be precursor mRNAs. There are four families of tasiRNA-producing loci (*TAS* genes) in *Arabidopsis*, categorized into two types based on differences in biogenesis, known as "one-hit" and "two-hit." In the "one-hit" model, a single targeting site at the 5′-end of the *TAS1*, *TAS2*, or *TAS4* precursor is targeted and sliced by a 22-nt AGO1-bound-miRNA, converted into dsRNA, and diced in a precise head-to-tail arrangement from 5′ to 3′ by DCL4. In the "two-hit" model of *TAS3*, two target sites for miRNA-bound AGO7 are required for the processing of *TAS3* precursors, but only the target site on the 3′-end is cleaved, with the tasiRNAs generated in the 3′ to 5′ direction. A specific *TAS3*-derived tasiRNA, aka tasiARF, targets *AUXIN RESPONSE FACTOR* (*ARF*) genes. Mis-regulation of tasiARF results in variable developmental defects in several plant species, such as *Arabidopsis* (Hunter et al. 2006), rice (Liu et al. 2007), maize (Dotto et al. 2014), and tomato (Yifhar et al. 2012). The *TAS3* tasiARF is highly conserved across land plants, and it has coevolved with the ARF target transcripts (Xia et al. 2017).

tasiRNAs are so named because tasiARF functions in *trans*; thus, the name emphasizes this aspect of their function. Yet, many other phased siRNAs are produced from both lncRNA "*TAS*" loci, and from protein-coding gene families, such as those encoding nucleotide-binding, leucine-rich repeat (NB-LRR)-encoding disease resistance genes, myeloblastosis (MYB) transcription factors, and other genes. Thus, tasiRNAs are one subset of the larger group of loci producing phasiRNAs. Although there are only eight lncRNA-derived *TAS* loci in *Arabidopsis*, there are >25 protein-coding sources of phasiRNAs (Howell et al. 2007); in other plant species, hundreds or even thousands of phasiRNA-generating loci have been identified. One particularly interesting group are the abundant reproductive phasiRNAs that have been found in maize and rice, highly enriched in anthers. These reproductive phasiRNAs fall into two classes: miR2118-triggered 21-nt phasiRNAs, enriched in pre-meiotic anther stages, and miR2275-triggered 24-nt phasiRNAs enriched in meiotic anthers (Johnson et al. 2009; Zhai et al. 2015). The accumulation patterns of these reproductive phasiRNAs in very specific stages suggest a critical role of phasiRNAs during the anther developmental process, but more studies are still required to characterize exactly what that role is.

8.4.2 Biogenesis of phasiRNAs

The production of reproductive phasiRNAs starts with mRNA precursors generated by RNA polymerase II (Pol II), yielding mature, polyadenylated mRNAs that are believed to be noncoding. The "one-hit" model describes most cases of phasiRNA biogenesis in which AGO utilizes a 22-nt miRNA to mediate the slicing of a target. MicroRNAs that are 22-nt (as opposed to most 21-nt miRNAs) typically facilitate the recruitment to targets of the machinery for phasiRNA biogenesis, namely RDR6 and SUPPRESSOR OF GENE SILENCING 3 (SGS3); these proteins convert the single-stranded mRNA precursor into dsRNA. DCL4 then recognizes the dsRNA, perhaps via recruitment by AGO1 or other components, and cleaves or "dices" the dsRNA into 21-nt phasiRNAs, working in the direction from 5′ to 3′, downstream of the cleaved target site. Following the conversion to dsRNA by RDR6/SGS3, DCL4 again cleaves the dsRNA into 21-nt tasiRNAs, but operating from the 3′ end of the precursor to the 5′-end (in the direction upstream of the cleaved target site). This biogenesis process takes place in the cytosol and requires polysome loading (Li et al. 2016).

Of the four canonical Dicer proteins encoded in a typical eudicot genome (like *Arabidopsis*), two are involved in phasiRNA biogenesis. This includes DCL1 for its activity in producing

miRNA triggers (see Sect. 5), and DCL4 for its activity in producing the 21-nt phasiRNAs. However, some monocots have evolved a new Dicer family member, known as DCL5, which emerged from DCL3 at some as-yet-unknown point in the diversification of monocots (Margis et al. 2006). DCL5, formerly known as DCL3b, is the Dicer enzyme that produces the meiotic 24-nt reproductive phasiRNA in grass anthers (Song et al. 2012). Future work may reveal where and how in the monocot lineage this gene duplication event yielded a novel participant in the production of phasiRNAs.

8.4.3 The Function of Reproductive phasiRNAs

Though reproductive phasiRNAs in maize are quite diverse and numerous, their functional roles are not well elucidated by experimental analysis. However, one can speculate as to their function based on the activities of plant sRNAs of similar lengths; in other words, 21-nt tasiRNAs may function in posttranscriptional control of targets, while 24-nt hc-siRNAs may direct chromatin modifications at their target loci. In each case, sRNAs could be incorporated into the RNA-induced silencing complex (RISC), with sequence homology directing the silencing of their corresponding targets. Bioinformatics studies have not yet described targets other than in *cis*—the *PHAS* loci themselves, so one possibility is a self-regulatory circuit. One such analysis in rice panicle sRNAs has illustrated the molecular basis that 21-nt phasiRNAs can target *in cis* to regulate their own precursors (Tamim et al. 2018). The observation of elevated DNA methylation at *PHAS* loci during zygotene in maize anthers provides evidence for this hypothesis (Dukowic-Schulze et al. 2017). In this context, reproductive 21- and 24-nt phasiRNAs may be functionally analogous to miRNAs in fission yeast or Piwi-interacting RNA (piRNA) in animals—both of which are specific to the process of meiosis. More experimental evidence will be needed to validate this theory. With the advent of site-directed mutation enabled by the technology called clustered regularly interspaced short palindromic repeats (CRISPR), it is possible to knock out the key components of the biogenesis pathway to study alterations such as to chromatin structure during meiosis in the absence of phasiRNA activity.

8.5 miRNAs

8.5.1 Overview of Plant miRNAs

MicroRNAs (miRNAs) are a class of small noncoding RNAs in eukaryotes that regulate gene expression via posttranscriptional gene silencing. In plants, miRNAs have been identified to regulate key processes such as development, growth, and stress response (Budak and Akpinar 2015). Plant miRNA biogenesis and miRNA-induced silencing have been studied extensively in *Arabidopsis*, and the biogenesis pathway is well conserved in all other plants that have been examined. This biogenesis begins with the transcription of noncoding RNAs by Pol II from miRNA genes (*MIRNAs*) that were often not annotated in early versions of many plant genomes, since gene-finding software may fail to find lncRNAs. Once a *MIRNA* gene is transcribed by Pol II, the resulting transcript matures as a typical mRNA, with addition of a 5′ 7-methylguanosine cap and 3′ polyadenylated tail. This product, deemed a primary miRNA transcript (pri-miRNA), folds back onto itself to create the classical hairpin structure of a pri-miRNA. This pri-miRNA is then recognized by DCL1, a member of the Dicer-like family of enzymes, perhaps by recruitment by other protein partners (Budak and Akpinar 2015; Rogers and Chen 2013). Through structural features of the pri-miRNA, DCL1 can identify a location near one end of the pri-miRNA for cleavage into a precursor miRNA (pre-miRNA) (Budak and Akpinar 2015). DCL1 carries out a second and final cleavage of the pre-miRNA to produce the miRNA-miRNA* duplex, a short dsRNA in which one strand produces the mature miRNA guide and the other produces the passenger miRNA (miRNA*). This dsRNA is characterized

by a 2-nucleotide 3′ overhang on both strands (Budak and Akpinar 2015).

The processing of miRNAs takes place in the nucleus; however, mature miRNAs must be transported to the cytoplasm prior to functioning in their role in the suppression of protein production from target mRNAs. This transfer of miRNA from the nucleus to the cytoplasm is not fully understood yet. However, data have shown that HASTY (HST), an exportin-related protein, functions as a key enzyme involved in this step (Park et al. 2005), though it has been observed that some miRNAs can still accumulate in the cytoplasm in *hst* mutants. To protect the miRNA-miRNA* duplex from degradation via uridylation, it is stabilized with a 3′ terminal methyl group by HEN1. It is unclear if this occurs in the nucleus or the cytoplasm (Budak and Akpinar 2015; Park et al. 2005). Once in the cytoplasm, the mature miRNA guide separates from the passenger miRNA and is loaded into an Argonaute protein, typically AGO1, to form the RISC. This RISC, which is comprised of numerous proteins, uses the mature miRNA guide as a template to identify targets to either cleave a mRNA or facilitate translational inhibition (Budak and Akpinar 2015). Targets of miRNAs are identified by their near perfect complementarity to the miRNA. When suppressing a target via mRNA cleavage, the target site of a miRNA is found on the mRNA, while the miRNA is still loaded in the RISC. In plants, cleavage is typically observed between the 10th and 11th positions of the alignment and is facilitated by AGO1 (Huntzinger and Izaurralde 2011). Products of this cleavage are detectable using parallel analysis of RNA ends (PARE) libraries (German et al. 2009). Combining these PARE data with miRNA-target prediction tools, such as sPARTA (Kakrana et al. 2014) and CleaveLand (Addo-Quaye et al. 2009), can validate predicted targets of miRNAs on a genome-wide basis.

Upon miRNA-directed cleavage of a target mRNA, the cleaved mRNA molecule is no longer competent to produce protein products. Disrupting a miRNA would therefore typically result in the overaccumulation of miRNA-target gene products, relative to wild-type controls. This phenomenon is visible in numerous maize mutants, some of which were identified long before miRNAs were described. In many plants, miRNAs exist in large numbers (hundreds of distinct loci) and can regulate the production of numerous proteins. In the following paragraphs, we explore a few cases of maize miRNAs and targets that have been particularly well characterized.

8.5.2 Mutations in miRNA Biogenesis and Silencing Pathways

8.5.2.1 Fuzzy Tassel

The *fuzzy tassel* (*fzt*) phenotype was identified in EMS-mutagenized plants, and it was later shown to result from a mutation in the maize ortholog of *DCL1* (Thompson et al. 2014). The *fzt* mutant plants have numerous developmental and vegetative defects, including the following: reduced plant stature, reduced number of leaves, missing and shorter internodes, reduced leaf surface area, and complete sterility in both ear and tassel due to various developmental defects in both tissues (Thompson et al. 2014). This mutation was mapped, cloned, and sequenced, showing that the *fzt* mutation is a single-base-pair mutation in the RNase IIIa domain of *DCL1*, a domain critical in the cleavage of both pri-miRNA and pre-miRNAs (Thompson et al. 2014). The differential expression analysis of miRNA abundance and their targets supports the hypothesis that the pleiotropic effects of this mutation are caused by the impacted miRNA abundances. Additionally, it was observed that not all known miRNAs had their abundances reduced; it appears that some miRNAs are only moderately affected, while others are entirely unaffected (Thompson et al. 2014). The basis of the differential impact on miRNAs is unclear, but could result from variation in how the altered DCL1 protein interacts with different precursors.

8.5.2.2 Corngrass1

Maize *Corngrass1* (*Cg1*) was a mutant first identified in 1947 and named for its closer

resemblance to other grasses than to wild-type maize (Whaley and Leech 1950). This mutation, deemed a "macromutation" for its profound effect in making the organism unrecognizable as belonging to its species (Chuck et al. 2007), results in the prolonged juvenile stage of maize plants. During this transition phase of plant development, called heteroblasty, there are numerous changes in leaf morphology and meristem fate. In 2007, researchers from Sarah Hake's group identified that the *Cg1* mutant encodes a tandem duplicate of miR156 genes, *zma-miR156b* and *zma-miR156c*, and the defects in the mutant are due to an overexpression of these miRNAs (Chuck et al. 2007). An analysis of transcript levels was conducted to investigate the expression levels of known targets of zma-miR156 and seven of the 13 predicted target genes, including *tga1* a gene known to have played a role in the domestication of maize from teosinte, were found to be downregulated in *Cg1* (Chuck et al. 2007).

8.5.2.3 Glossy15

Similar to the *Corngrass1* mutation, the *Glossy15* (*Gl15*) mutation extends the duration of the juvenile-to-adult phase transition in maize; this occurs via the overexpression of the *gl15* gene (Moose and Sisco 1994). Work done in *Arabidopsis* (Aukerman and Sakai 2003; Chen 2004) demonstrated that *miR172* can suppress the activity of *APETELA2*, the *Arabidopsis* ortholog of *Gl15*. The *APETELA2* target sequence of *miR172* is present in the *gl15* transcript, and thus, it was hypothesized that a *miR172* homolog could be responsible for the transition from juvenile-to-adult vegetative development in maize. Like *Arabidopsis*, *miR172* was found to not be present in early shoot development; it is only during the transition from juvenile to adult that *miR172* is detected, coincident with a decline in *gl15* mRNA levels. By increasing *gl15* activity, vegetative and reproductive phase changes are delayed, while *miR172* accumulation is not substantially different from wild type. It was later found that the *gl15* mutation contained a mutation in the target site of miR172. This mutation prevents the negative repression of *GL15* transcripts that occurs in wild-type maize during the juvenile-to-adult transition (Lauter et al. 2005). Thus, *miR172* promotes the transition to reproductive development by restricting the accumulation of GL15, but such a transition can only occur at a threshold of opposing activity of GL15 and *zma-miR172* (Aukerman and Sakai 2003).

8.5.3 Overview of Annotated Maize miRNAs

miRBase (http://www.mirbase.org/) is a public miRNA sequence repository designed to store all published miRNA sequences and their associated annotations and assign consistent identifiers to each published miRNA. As of version 22 (released March, 2018), there were 207 distinct mature miRNAs identified in maize being derived from 173 *MIR* genes, and all 203 miRNAs belonged to one of 31 miRNA families (Dezulian et al. 2005; Johnson et al. 2009; Maher et al. 2004; Thieme et al. 2011; Zhang et al. 2009). Via the use of next-generation sequencing, novel bioinformatics tools, and unique mutants, the identification of new miRNAs is becoming increasingly easier. Such use of these tools has been previously utilized in rice in which 76 new mature miRNA sequences were identified using 62 sRNA libraries (Jeong et al. 2011).

With a centralized miRNA repository like miRBase, it is easy to access and identify miRNAs from diverse species and identify similarities between them. Many miRNA families have been found to be conserved across numerous plant species. Some miRNAs are ancient with origins dating back to mosses, but are still identifiable in any plant due to their fundamental roles in the regulation of transcription that is required in the basics of plant life (Chávez Montes et al. 2014). Other miRNAs are a lot newer and emerged only within distinct branches of the angiosperms. By tracking the presence of miRNAs in a phylogenetic context, their evolutionary emergence can be discerned. The conserved and divergence of target sequences is also relevant to the study of miRNAs, as evidence of

conservation identifies important functional roles that have been selectively maintained, while divergence can yield new regulatory pathways (Axtell 2013).

8.6 Bioinformatics Tools

The sequencing of sRNAs produces a tremendous amount of data that requires comprehensive bioinformatics tools for proper analysis. There are numerous tools capable of assisting in the analysis of such data, but here we will only discuss a few, focusing on our tools as they are most familiar to us, particularly in the context of maize analyses. We recognize this bias and recognize that there are many other tools that are available for similar analyses. We have tried to list alternatives to each tool as there are many viable options to completing these analyses.

8.6.1 Shortstack

Shortstack (Shahid and Axtell 2014) is a comprehensive analytical tool developed by the laboratory of Michael Axtell with the purpose of analyzing mapped sRNA sequencing data. This tool is a general-purpose analytical tool to classify sRNAs. It requires as input little more than the sRNA data itself and a genome to which to map the sRNAs. In six steps, *Shortstack* works to annotate and quantify sRNA data generated to serve as the input file. These six steps include the following:

a. De novo discovery of clusters of sRNAs.
b. Quantification and phasing analysis.
c. Retrieval of genomic sequences for RNA folding.
d. Identification of hairpin structures that may qualify as miRNA precursors.
e. Annotation of hairpin association and miRNA candidates.
f. Output of organized and summarized results.

Within each of these steps, *Shortstack* makes use of the known biology of various classes of sRNAs to characterize genes producing sRNAs, sRNA clusters, miRNA genes, and phasiRNAs (Axtell 2013). While running *Shortstack* is generally quite simple, there are many options that may be modified by the user, so it is highly recommended to read all the information provided within the README and even try utilizing some of the test data prior to running new analyses.

8.6.2 sPARTA

sPARTA (Kakrana et al. 2014) is a tool developed by the Meyers group for the purpose of rapidly and accurately predicting and validating targets of miRNAs. As discussed previously, miRNAs function to suppress the expression of a gene by matching with high sequence complementarity to an mRNA. When the mRNA target of a miRNA is identified, it is cleaved by AGO1. This cleavage is identifiable via the sequencing of PARE data, mentioned above. By predicting the locations at which a miRNA directs cleavage and by identifying a PARE signal at that same location, cleavage of an mRNA facilitated by a miRNA can be validated. *sPARTA* first predicts targets of miRNAs and then utilizes PARE libraries to validate these miRNA-target interaction (Kakrana et al. 2014). Several alternatives to *sPARTA* have been described, including *CleaveLand* (Addo-Quaye et al. 2009), *PAREsnip* (Folkes et al. 2012), and *SeqTar* (Zheng et al. 2012).

8.6.3 PHASIS

PHASIS is a suite of tools that functions to rapidly identify *PHAS* loci in plants. This tool, also developed in the Meyers group, has three separate components to discover, annotate, and quantify *PHAS* loci, as well as identify the miRNA triggers for those *PHAS* loci. These components—phasdetect, phasmerge, and phastrigs—each have distinct purposes in the pipeline and were separated to allow users some flexibility in the inputs provided to each. While not

required, utilizing PARE data in a *PHASIS* run will allow for experimental validation of predicted cleavage locations, assisting with the identification of miRNA triggers (Kakrana et al. 2017). Some alternatives to *PHASIS* are *PhaseTank* (Guo et al. 2015) and *unitas* (Gebert et al. 2017).

8.6.4 Plant Small RNA Similarity Tool

This tool was developed with the purpose of identifying similar miRNAs across species to detect the conservation of miRNAs across genomes. It uses an implementation of BLAST, and it has evolved to identify similar sRNAs regardless of the class of sRNA. Thus, using a list of user-provided sequences, it is capable of searching for the presence of related sequences in over 36 public databases hosted by the Meyers lab (https://mpss.danforthcenter.org). This tool's primary use is to identify the presence and abundance of one or more sRNAs and close variants in one or more library from another plant species. It allows for rapid identification of homologs of mature miRNAs, for example.

8.6.5 miRBase

As previously mentioned, miRbase is a public miRNA sequence repository designed to store all published miRNA sequences and their associated annotations, as well as assign consistent identifiers to each published miRNA. Consistency among identified miRNA names allows comparisons across species. Despite the great utility of miRBase, it is not without issues. For example, version 21 was published in June 2014 and has been untouched and unrevised for over three and a half years, as of our writing of this chapter. During this time, many miRNAs were identified and submitted to miRBase, and even published in peer-reviewed articles, but the lack of updates kept those newly identified miRNAs unnamed or unassigned, and the data on miRNAs are not yet in the public domain.

8.7 Conclusions

Research into maize small RNAs has advanced rapidly over the last decade. This has resulted in the identification of numerous classes of sRNAs, the populations of sRNAs (and their genomic source loci) that comprise those classes, and the functions of many of these sRNAs. Functional analysis has described unique aspects of maize sRNA biology, providing a motivation for further, detailed molecular, genetic and biochemical studies. Many questions remain, for example, what are the roles of sRNAs in paramutation, and why is this phenomenon so much more evident in maize than other species? What are the functions of reproductive phasiRNAs in maize and other grasses? Generally speaking, how often do lineage-specific miRNAs arise in plants, and are such examples limited to maize and its closest relatives? What are the functions of the AGO proteins in maize, a set nearly twice as numerous as those of *Arabidopsis*? How do other small RNA effector and biogenesis proteins compare in their functions, from the grasses across the angiosperms? These and many other exciting questions are sure to drive research into maize small RNA biology for the coming decade.

Funding This work was supported by a grant from the US National Science Foundation, Plant Genome Research Program (NSF-PGRP), award #1649424.

References

Addo-Quaye C, Miller W, Axtell MJ (2009) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics 25(1):130–131. https://doi.org/10.1093/bioinformatics/btn604

Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, Sikkink K, Chandler VL (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442 (7100):295–298. https://doi.org/10.1038/nature04884

Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K (2010) RNA-mediated *trans*-communication can establish paramutation at the *b1*

locus in maize. Proc Natl Acad Sci U S A 107 (29):12986–12991

Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. Plant Cell 15(11):2730–2741. https://doi.org/10.1105/tpc.016238.pression

Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64:137–159. https://doi.org/10.1146/annurev-arplant-050312-120043

Barbour J-ER, Liao IT, Stonaker JL, Lim JP, Lee CC, Parkinson SE, Kermicle J, Simon SA, Meyers BC, Williams-Carrier R, Barkan A, Hollick JB (2012) Required to maintain repression2 is a novel protein that facilitates locus-specific paramutation in maize. Plant Cell 24(5):1761–1775

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol 16 (12):727–741. https://doi.org/10.1038/nrm4085

Brink RA (1956) A genetic change associated with the *R* locus in maize which is directed and potentially reversible. Genetics 41:872–889

Budak H, Akpinar BA (2015) Plant miRNAs: biogenesis, organization and origins. Funct Integr Genomics 15 (5):523–531. https://doi.org/10.1007/s10142-015-0451-2

Chávez Montes RA, Rosas-Cárdenas dFF, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, Marsch-Martínez N, Meyers BC, Green PJ, de Folter S (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun 5:3722

Chen X (2004) A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. Science 303(5666):2022–2025. https://doi.org/10.1126/science.1088060.A

Chuck G, Cigan AM, Saeteurn K, Hake S (2007) The heterochronic maize mutant *Corngrass1* results from overexpression of a tandem microRNA. Nat Genet 39 (4):544–549. https://doi.org/10.1038/ng2001

Dezulian T, Palatnik JF, Huson D, Weigel D (2005) Conservation and divergence of microRNA families in plants. Genome Biol 6(11):13. https://doi.org/10.1186/gb-2005-6-11-p13

Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL (2000) *mediator of paramutation1* is required for establishment and maintenance of paramutation at multiple maize loci. Plant Cell 12(11):2101–2118. https://doi.org/10.1105/tpc.12.11.2101

Dotto MC, Petsch KA, Aukerman MJ, Beatty M, Hammell M, Timmermans MCP (2014) Genome-wide analysis of *leafbladeless1*-regulated and phased small RNAs underscores the importance of the *TAS3* ta-siRNA pathway to maize development. PLoS Genet 10(12):e1004826. https://doi.org/10.1371/journal.pgen.1004826

Dukowic-Schulze S, Liu C, Chen C (2017) Not just gene expression: 3D implications of chromatin modifications during sexual plant reproduction. Plant Cell Rep 37(1):1–6. https://doi.org/10.1007/s00299-017-2222-0

Erhard KFJ, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB (2009) RNA polymerase IV functions in paramutation in *Zea mays*. Science 323(5918):1201–1206

Fei Q, Xia R, Meyers BC (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. Plant Cell 25(7):2400–2415. https://doi.org/10.1105/tpc.113.114652

Folkes L, Moxon S, Woolfenden HC, Stocks MB, Szittya G, Dalmay T, Moulton V (2012) PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. Nucleic Acids Res 40(13):e103. https://doi.org/10.1093/nar/gks277

Gabriel JM, Narain A, Talbot J-ERB, Liao IT, Simon SA, Kong G, Meyers BC, Hollick JB (2015) *dicer-like3* is required for paramutation, small RNA biogenesis, and normal development in *Zea mays*. Poster presented at the Maize Genetics Conference

Gebert D, Hewel C, Rosenkranz D (2017) Unitas: the universal tool for annotation of small RNAs. BMC Genom 18(1):644. https://doi.org/10.1186/s12864-017-4031-9

German MA, Luo S, Schroth G, Meyers BC, Green PJ (2009) Construction of parallel analysis of RNA ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc 4(3):356–362

Guo Q, Qu X, Jin W (2015) PhaseTank: genome-wide computational identification of phasiRNAs and their regulatory cascades. Bioinformatics 31(2):284–286. https://doi.org/10.1093/bioinformatics/btu628

Hale CJ, Erhard KF, Lisch D, Hollick JB (2009) Production and processing of siRNA precursor transcripts from the highly repetitive maize genome. PLoS Genet 5(8):e1000598. https://doi.org/10.1371/journal.pgen.1000598

Hale CJ, Stonaker JL, Gross SM, Hollick JB (2007) A novel Snf2 protein maintains *trans*-generational regulatory states established by paramutation in maize. PLoS Biol 5(10):e275

Hollick JB (2016) Paramutation and related phenomena in diverse species. Nat Rev Genet 18(1):5–23. https://doi.org/10.1038/nrg.2016.115

Hollick JB, Chandler VL (2001) Genetic factors required to maintain repression of a paramutagenic maize *pl1* allele. Genetics 157(1):369–378

Hollick JB, Kermicle JL, Parkinson SE (2005) *Rmr6* maintains meiotic inheritance of paramutant states in *Zea mays*. Genetics 171(2):725–740. https://doi.org/10.1534/genetics.105.045260

Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC (2007) Genome-wide analysis of the nPOLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell 19(3):926–942. https://doi.org/10.1105/tpc.107.050062

Hunter C, Willmann MR, Wu G, Yoshikawa M, MdlL Gutirrez-Nava, Poethig SR (2006) *Trans*-acting

siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. Development 133(15):2973–2981. https://doi.org/10.1242/dev.02491

Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 12(2):99–110

Jeong D-H, Park S, Zhai J, Gurazada SGR, De Paoli E, Meyers BC, Green PJ (2011) Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. Plant Cell 23(12):4185

Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, Sundaresan V, Vance V, Bowman LH (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res 19(8):1429–1440. https://doi.org/10.1101/gr.089854.108

Kakrana A, Hammond R, Patel P, Nakano M, Meyers BC (2014) *sPARTA*: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. Nucleic Acids Res 42(18):2973–2981. https://doi.org/10.1093/nar/gku693

Kakrana A, Li P, Patel P, Hammond R, Anand D, Mathioni S, Meyers B (2017) PHASIS: a computational suite for de novo discovery and characterization of phased, siRNA-generating loci and their miRNA triggers. bioRxiv

Lauter N, Kampani A, Carlson S, Goebel M, Moose SP (2005) *microRNA172* down-regulates *glossy15* to promote vegetative phase change in maize. Proc Natl Acad Sci U S A 102(26):9412–9417

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204–220. https://doi.org/10.1038/nrg2719

Li S, Le B, Ma X, Li S, You C, Yu Y, Zhang B, Liu L, Gao L, Shi T, Zhao Y, Mo B, Cao X, Chen X (2016) Biogenesis of phased siRNAs on membrane-bound polysomes in *Arabidopsis*. eLife 5:e22750. https://doi.org/10.7554/elife.22750

Liu B, Chen Z, Song X, Liu C, Cui X, Zhao X, Fang J, Xu W, Zhang H, Wang X, Chu C, Deng X, Xue Y, Cao X (2007) *Oryza sativa Dicer-like4* reveals a key role for small interfering RNA silencing in plant development. Plant Cell 19(9):2705–2718. https://doi.org/10.1105/tpc.107.052209

Luo J, Hall BD (2007) A multistep process gave rise to RNA polymerase IV of land plants. J Mol Evol 64 (1):101–112

Maher C, Timmermans M, Stein L, Ware D Identifying microRNAs in plant genomes. In: Proceedings IEEE Computational Systems Bioinformatics Conference, 2004, pp 718–723

Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM (2006) The evolution and diversification of dicers in plants. FEBS Lett 580(10):2442–2450. https://doi.org/10.1016/j.febslet.2006.03.072

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15(6):394–408. https://doi.org/10.1038/nrg3683

Moose SP, Sisco PH (1994) *Glossy15* controls the epidermal juvenile-to-adult phase transition in maize. Plant Cell 6(10):1343–1355. https://doi.org/10.1105/tpc.6.10.1343

Nobuta K, Lu C, Shrivastava R, Ma Pillay (2008) Distinct size distribution of endogeneous siRNAs in maize: evidence from deep sequencing in the *mop1-1* mutant. Proc Natl Acad Sci U S A 105(39):14958–14963. https://doi.org/10.1073/pnas.0808066105

Park MY, Wu G, Gonzalez-sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in *Arabidopsis*. Proc Natl Acad Sci U S A 102(10):3691–3696

Pikaard CS, Haag JR, Ream T, Wierzbicki AT (2008) Roles of RNA polymerase IV in gene silencing. Trends Plant Sci 13(7):390–397

Pikaard CS, Tucker S (2009) RNA-silencing enzymes Pol IV and Pol V in maize: more than one flavor? PLoS Genet 5(11):e1000736. https://doi.org/10.1371/journal.pgen.1000736

Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell 25(7):2383–2399. https://doi.org/10.1105/tpc.113.113159

Shahid S, Axtell MJ (2014) Identification and annotation of small RNA genes using shortstack. Methods 67 (1):20–27

Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Vyas M, Kermicle J, Jurcin D, Brzeski J, Cai Y, Chandler VL (2009) A dominant mutation in *mediator of paramutation2*, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. PLoS Genet 5(11): e1000725. https://doi.org/10.1371/journal.pgen.1000725

Sloan AE, Sidorenko L, McGinnis KM (2014) Diverse gene-silencing mechanisms with distinct requirements for RNA polymerase subunits in *Zea mays*. Genetics 198(3):1031–1042

Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong DH, Nakano M, Cao S, Liu C, Chu C, Wang XJ, Green PJ, Meyers BC, Cao X (2012) Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. Plant J 69 (3):462–474. https://doi.org/10.1111/j.1365-313X.2011.04805.x

Stonaker JL, Lim JP, Erhard KF, Hollick JB (2009) Diversity of Pol IV function is defined by mutations at the maize *rmr7* locus. PLoS Genet 5(11):e1000706. https://doi.org/10.1371/journal.pgen.1000706

Tamim S, Cai Z, Mathioni S, Zhai J, Teng C, Zhang Q, Meyers BC (2018) The basis of accumulation differences in plant 21-nt reproductive phasiRNAs, and their *cis*-directed activity. bioRxiv

Thieme CJ, Gramzow L, Lobbes D, Theißen G (2011) SplamiR—prediction of spliced miRNAs in plants. Bioinformatics 27(9):1215–1223. https://doi.org/10.1093/bioinformatics/btr132

Thompson BE, Basham C, Hammond R, Ding Q, Kakrana A, Lee T-F, Simon SA, Meeley R, Meyers BC, Hake S (2014) The *dicer-like1* homolog *fuzzy tassel* is required for the regulation of meristem determinacy in the inflorescence and vegetative growth in maize. Plant Cell 26(12):4702–4717. https://doi.org/10.1105/tpc.114.132670

Wang F, Axtell MJ (2017) *AGO4* is specifically required for heterochromatic siRNA accumulation at Pol V-dependent loci in *Arabidopsis thaliana*. Plant J 136(4):669–687. https://doi.org/10.1111/tpj.13463

Whaley WG, Leech JH (1950) The developmental morphology of the mutant "Corn Grass". Bull Torrey Bot Club 77(4):274–286

Xia R, Xu J, Meyers BC (2017) The emergence, evolution, and diversification of the *miR390-TAS3-ARF* pathway in land plants. Plant Cell 29(6):1232–1247. https://doi.org/10.1105/tpc.17.00185

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biol 2(5):e104

Yifhar T, Pekker I, Peled D, Friedlander G, Pistunov A, Sabban M, Wachsman G, Alvarez JP, Amsellem Z, Eshed Y (2012) Failure of the tomato *trans*-acting short interfering RNA program to regulate AUXIN RESPONSE FACTOR3 and ARF4 underlies the wiry leaf syndrome. Plant Cell 24(9):3575–3589. https://doi.org/10.1105/tpc.112.100222

Zhai J, Zhang H, Arikit S, Huang K, Nan G-L, Walbot V, Meyers BC (2015) Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. Proc Natl Acad Sci U S A 112(10):3146–3151. https://doi.org/10.1073/pnas.1418918112

Zhang L, Chia J-M, Kumari S, Stein JC, Liu Z, Narechania A, Maher CA, Guill K, McMullen MD, Ware D (2009) A genome-wide characterization of microRNA genes in maize. PLoS Genet 5(11):e1000716

Zheng Y, Li YF, Sunkar R, Zhang W (2012) SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. Nucleic Acids Res 40(4):e28. https://doi.org/10.1093/nar/gkr1092

Part III
Genomic and Germplasm Resources

9 The UniformMu Resource: Construction, Applications, and Opportunities

Donald R. McCarty, Peng Liu and Karen E. Koch

Abstract

Invaluable insights into functional genomics have arisen from knockout and knockdown mutants generated by transposon mutagenesis. Thousands of insertional mutants are available free of charge from the UniformMu national public resource for maize. This resource was created using the native Robertson's Mutator system and resulting features include (1) an inbred genetic background ideal for phenotype analysis; (2) effective genetic control of Mu transposon activity that facilitates genetic and molecular analyses; (3) precise mapping of transposon insertions enabled by targeted sequencing (Mu flanks identified by a Mu-seq protocol for high-throughput genotyping); (4) cost-free, worldwide distribution of high-quality, sustainable seed stocks through MaizeGDB.org and the Maize Genetics Cooperation Stock Center. Available materials have been especially useful for genetic analysis of complex, multi-genetic traits such as domestication, seed development, and disease resistance. Additional applications for UniformMu resources include new strategies for both forward- and reverse-genetics (phenotype-to-genotype or the reverse) as well as synergies with emerging gene-editing technologies (e.g., MuCRISPR).

D. R. McCarty (✉) · P. Liu · K. E. Koch
Horticultural Sciences Department, Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL, USA
e-mail: drm@ufl.edu

P. Liu
e-mail: mcliup@ufl.edu

K. E. Koch
e-mail: kekoch@ufl.edu

D. R. McCarty · K. E. Koch
Genetics Institute, University of Florida, Gainesville, FL, USA

9.1 Introduction

9.1.1 Overview of Insertional-Mutant Resources

Loss-of-function mutants are invaluable tools for linking genes and networks to phenotypes. Confirmation of a causal role between genotype and phenotype is central to establishing a baseline for interpreting the biological role of a gene. The relationship provides a foundation for hypotheses and further experimental analysis. Genetics researchers thus dream of having "finger-tip" access to loss-of-function mutants for every gene in a genome of interest. Also, to facilitate quantitative analysis of phenotypes, all mutants should ideally be available in a common, well-defined, preferably inbred genetic background. In recent decades, high-throughput insertional mutagenesis in such backgrounds

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_9

has played a key role in systematic functional analysis of plant genomes (Alonso et al. 2003; Hirochika et al. 2004; McCarty et al. 2005, 2013a). In Arabidopsis, for example, readily available, sequence-indexed T-DNA insertion lines provide nearly comprehensive coverage of the genome. The extent of this coverage has greatly facilitated genetic analyses of diverse contributions to development, metabolism, and responses to abiotic and biotic stresses (Alonso et al. 2003). Similar resources, though somewhat less comprehensive, have been developed for other model organisms including maize (Jeon et al. 2000; May et al. 2003; McCarty et al. 2005; Vollbrecht et al. 2010; Williams-Carrier et al. 2010; Walbot and Qüesta 2012; Li et al. 2013). Other strategies for generating comprehensive collections of knockout mutants have been employed or contemplated and include TILLING (McCallum et al. 2000), radiation-induced deletions (Shirley et al. 1992), and CRISPR/Cas9 (Shan et al. 2013; Belhaj et al. 2015).

9.1.2 Transposon Resources in the Age of Gene-Editing

Although CRISPR/Cas9 is an especially promising technology, insertional mutagenesis offers some complementary advantages as well as opportunities for synergy. In principle, CRISPR-Cas9 methods offer capabilities that have greater reach and specificity than conventional insertional mutagenesis resources, especially when focused on small numbers of genes. For such applications, CRISPR/Cas9 may indeed eventually supplant insertion-based resources for reverse-genetics (Belhaj et al. 2015). However, considerations entailed in such an assessment are complex and not limited to scientific issues. The current high cost and relative inefficiency of methods for transformation of maize inbreds are key factors that will likely impact genome-wide applications of CRISPR for the foreseeable future. In addition, since maize genetics experiments predominantly involve field-grown plants, a second practical consideration is whether plants carrying gene-edited alleles will be treated as transgenic organisms by regulatory authorities.

Over and above these considerations of cost and practicality, large collections of insertion mutants retain at least two key advantages over gene-editing approaches. One is the suitability of genome-wide mutant resources for forward-genetic strategies that link phenotypes to genotypes (O'Malley and Ecker 2010; Hunter et al. 2014). Second, an often under-appreciated advantage of mutations that share a common molecularly defined insertion is the feasibility of developing high-throughput genotyping platforms such as the Mu-targeted Mu-seq protocol (McCarty et al. 2013a).

When used in combination with emerging high-throughput phenotyping technologies (Fahlgren et al. 2015), high-throughput genotyping capabilities unlock an important synergy by enabling large-scale genetic analyses of insertion mutants to link genotypes to phenotypes. Thus, the genotyping methods that were initially developed to facilitate construction of large-scale genetic resources are now emerging as powerful approaches for bringing those resources to bear on genome-scale problems (O'Malley and Ecker 2010; Hunter et al. 2014). For similar reasons, transposon mutagenesis remains a popular strategy for functional analysis of diverse microbial genomes (Hayes 2003; Barquist et al. 2013; Niehaus et al. 2018).

9.1.3 Systematic Transposon Mutagenesis of Maize Genes

In maize, strategies for development of insertion-based functional genomics resources have focused on endogenous transposon systems. These systems have included *Activator/Dissociation* (*Ac/Ds*), *Robertson's Mutator* (*Mu*), and *Suppressor-Mutator* (*Spm*), with efforts utilizing engineered (Raizada et al. 2001; Li et al. 2013) as well as native elements (May et al. 2003; McCarty et al. 2005; Vollbrecht et al. 2010; Williams-Carrier et al. 2010; Walbot and Qüesta 2012). The widely deployed *Ac/Ds* and

Mutator transposon systems have distinctive behavioral characteristics that affect the choice of strategy used for large-scale mutagenesis. These distinctive reverse-genetics strategies are complementary in key respects.

Primary differences between *Ac/Ds* and *Mutator* systems that affect reverse-genetics strategies include (1) transposon copy number in active lines, (2) linkage bias of transpositions, and (3) insertion-site preferences in the maize genome. In contrast to *Mutator* (Lisch et al. 1995), *Ac/Ds* elements have a strong bias for transposition to nearby sites in the genome (Vollbrecht et al. 2010). For this reason, *Ac/Ds* elements have been employed for regional mutagenesis strategies that enable targeting of genes linked to an *Ac* or *Ds* transposon located at a previously mapped donor site (Brutnell and Conrad 2003; Vollbrecht et al. 2010). Regional mutagenesis can be especially effective for analyzing complex loci that include multiple members of a gene family (e.g., tandem duplications). Genome-wide coverage is obtained by establishing donor *Ds* insertions at many locations in the genome. Ideally, each donor *Ds* is maintained in a separate genetic stock giving the geneticist control over which region of the genome is to be targeted for mutagenesis (Brutnell and Conrad 2003; www.acdstagging.org). One can thus obtain relatively high mutation frequencies in the region of interest by concentrating the activity of a single *Ds* transposon in a local region. Although regional mutagenesis often requires multiple generations, results can be especially valuable for genes not targeted effectively by other systems.

An alternative strategy is provided by the high-copy-number *Mutator* system, which delivers a comparatively high mutation frequency over the entire genome. In contrast to *Ac/Ds,* the *Mutator* transposons exhibit little or no bias for transposition to linked sites in the genome (Lisch et al. 1995). The resulting high mutation frequency and breadth of coverage have led to wide use of *Mutator* for comprehensive mutagenesis of the maize genome (Bensen et al. 1995; May et al. 2003; McCarty et al. 2005; Williams-Carrier et al. 2010; Walbot and Qüesta 2012). The comparatively high rate of genome-wide mutation is attributable in part to the large number of active transposon copies per individual (typically ~50 *Mu's* each). This abundance of Mu copies initially posed challenges that were successfully addressed through development of sequence-based, high-throughput genotyping protocols suitable for molecular analysis of high-copy transposon lines (McCarty et al. 2013a).

Both *Mutator* and *Ac/Ds* transposons insert preferentially into genes rather than inter-genic regions rich in repetitive sequences (Cresse et al. 1995; Vollbrecht et al. 2010). This bias toward single-copy, non-repetitive regions of the genome increases the likelihood that each new transposition will cause a mutation by inserting in or near a gene. However, the two systems exhibit different positional biases within genes. The *Mutator* elements show a strong bias for insertion into a region of open chromatin near the transcription start site (Springer et al. 2018), whereas *Ds* insertions show a comparatively weak bias toward the 5′-end of genes (Vollbrecht et al. 2010).

To take advantage of these distinctive behavioral characteristics, maize geneticists have devised complementary strategies that leverage the respective copy number, linkage bias, and insertion-site preference characteristics of the *Mutator* and *Ac/Ds* systems.

Important broader challenges of using *Mutator* as the basis for a public reverse-genetics resource include (1) managing transposon copy number and genetic load in a population with a uniform genetic background conducive to quantitative analysis of phenotypes, (2) developing efficient methods for high-throughput mapping and genotyping of a high-copy number transposon in a large population, and (3) maintaining a high-quality, sustainable seed resource that is freely accessible to the research community and preserves the full-array of insertion mutants identified in the population. These challenges were addressed by incorporating into the UniformMu resource the six essential features described in the next section.

9.2 Construction of the UniformMu Resource

9.2.1 Key Design Goals of the UniformMu Resource

1. Construct a transposon population in an inbred genetic background that maximizes uniformity and thus enables detection and quantification of phenotypes linked to tagged genes.
2. Incorporate an efficient mechanism for genetic "on–off" control of *Mu* activity to facilitate molecular, genetic, and phenotype analyses of insertion mutants.
3. Manage the total number of transposed *Mu* elements carried by individuals in the population and prevent excessive accumulation of insertions in any given genome by using a steady-state transposon mutagenesis strategy (McCarty et al. 2005).
4. Minimize selection against deleterious mutations in the population by developing sustainable seed stocks that result from a sib-pollination strategy that preserves insertions in a heterozygous state.
5. Employ high-throughput, sequence-based transposon genotyping methods to efficiently identify and map germinal transposon insertions in thousands of carefully prepared seed stocks.
6. Provide public access to the resource using established community resource-portals that will ensure long-term sustainability of the delivery system.

9.2.2 Transposon Mutagenesis in a Phenotypically Uniform Inbred Background

Construction of a transposon population in an inbred genetic background maximizes the uniformity that facilitates detection and quantification of phenotypes linked to tagged genes. We chose the color-converted (ACR) derivative of the W22 inbred developed by Brink at the University of Wisconsin specifically for genetic studies (Bray and Brink 1966). Importantly, W22 (ACR) carries the genes for seed anthocyanin pigment that we needed to use as a genetic marker system in the "on–off" selection of transposase activity. Moreover, the same inbred has been used for development of key *Ac/Ds* resources (Vollbrecht et al. 2010) providing a common background for analysis of insertional mutants. To further enhance these collective efforts in maize, laboratories that spear-headed development of transposon resources also formed a consortium to sequence, annotate, and characterize the genome of W22 (ACR) (Springer et al. 2018; https://www.maizegdb.org/genome/genome_assembly/Zm-W22-REFERENCE-NRGENE-2.0).

9.2.3 Genetic Control of Mu Activity by Selection of Mu-on and Mu-off Plants

In order to study and maintain newly generated transposon mutants, each mutagenic line must be converted to a stable, non-mutagenic state. A genetically stable background is essential for (1) effective phenotypic comparisons between mutant and non-mutant individuals and (2) efficient molecular genotyping. More specifically, if a plant is "Mu-on," somatic insertions will arise from transpositions that occur in vegetative tissues during development (e.g., in leaf cells), and these can far out-number heritable insertions derived from the germline (lineage leading to pollen and egg nucleii). Leaf DNA sampled from a Mu-on plant may thus include thousands of somatic insertions that are not recoverable for genetic analysis. Importantly, this background of somatic insertions can complicate identification of germinal insertions in the genome of an individual, as well as subsequent high-throughput mapping and genotyping analyses needed to establish linkage of specific insertions with phenotypes. In addition, if continued Mu activity is not suppressed, it will lead to accumulation of new germinal insertions in lines selected for

genetic analysis. This increased genetic load may further confound phenotype analyses. A crucial feature for use of this high-copy-number system is therefore the capacity to activate and deactivate *Mutator* transposons.

Toward this end, we used the color-marker noted above because it allowed visualization and genetic selection of *Mu* transposase activity (Fig. 9.1). The *Bronze1* (*Bz1*) gene is required for synthesis of purple anthocyanin pigment in the aleurone cell layer of the endosperm. Recessive, loss-of-function *bz1* mutations confer a bronze-colored aleurone phenotype. The transposase that catalyzes movement of *Mu* insertion elements from place to place in the genome is encoded by *MuDR*, the so-called autonomous member of the *Mutator* family of transposable elements. Typically, the majority of *Mu* insertions are due to more numerous non-autonomous elements that do not have an intact transposase gene. The non-autonomous transposons can only move if at least one copy of *MuDR* is also present in the genome. The *bz1-mum9* allele carries a non-autonomous Mu-transposon insertion that can excise from the gene if *MuDR* is present elsewhere the genome. Thus, in the Mu-on plants, which carry *MuDR*, purple spots on bronze kernels result from frequent somatic reversion of the *bz1-mum9* allele to a functional wild-type (*Bz1*) gene (each spot is a clonal *Bz1* somatic sector). By contrast, in Mu-off plants that do not carry a *MuDR* transposase, *bz1-mum9* confers a stable bronze aleurone phenotype. In this way, the *bz1-mum9* allele provides a sensitive marker for genetic selection of Mu-on (spotted aleurone, *MuDR* present) and Mu-off (stable bronze, *MuDR* absent) plants. Typically, the Mu-on plants used to generate new transposon insertions in the maize genome carry one or two *MuDR* copies, plus 50 or more non-autonomous, but trans-activatable *Mu* elements. By enabling the presence and activity of *MuDR* to be visualized in segregating progeny of Mu-on plants, the

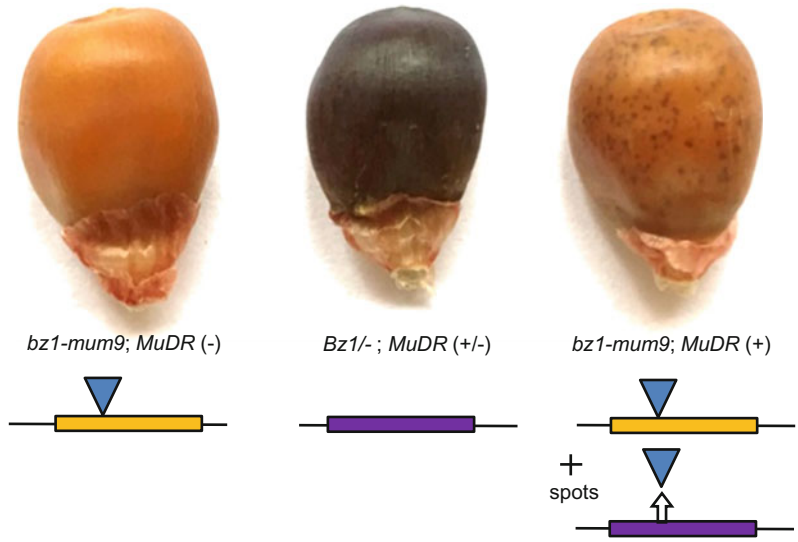

Fig. 9.1 Color marker for genetic control of Mu transposition. The bronze kernel mutant *bz1-mum9* carries a Mu transposon insertion in the *bronze1* (*bz1*) gene (long gold rectangle). The wild-type *Bz1* allele (purple rectangle) confers a purple-kernel phenotype in its homozygous or heterozygous state (middle kernel). The recessive *bz1-mum9* allele is inactivated by the *Mu* insertion. If there is not an autonomous *MuDR* transposon capable of expressing the transposase present somewhere in the genome *bz1-mum9* confers a stable bronze phenotype (left kernel) indicative of a heritable, Mu-off (*MuDR* (-)) state. When an autonomous *MuDR* transposon is present, the *Mu* transposon in *bz1-mum9* will excise frequently during endosperm development giving rise to multiple clonal sectors of aleurone cells that have restored *Bz1* function (bronze rectangle, lower right). These sectors which are indicative of the Mu-on (*MuDR* (+)) state are visible as purple spots (right kernel)

bz1-mum9 marker provides a fast and effective means for genetically controlling *Mu* transposition in the UniformMu population.

9.2.4 Management of Mutation Load and Forward Mutation Frequency

A steady-state mutagenesis strategy is used to manage genetic load by moderating copy number of transposed *Mu* elements in individual genomes. Steady-state mutagenesis (Fig. 9.2, McCarty et al. 2005) is achieved by a continuous backcrossing strategy that uses W22 (ACR) as the recurrent female parent, and *bz1-mum9* males that carry at least one copy of the autonomous *MuDR* transposon. Each Mu-on male plant is not only crossed to a W22 (ACR) female, but is also self-pollinated to allow analysis of its genotype. Evaluation of the ears from male plants is important to (1) confirm the presence of *MuDR* (densely spotted *bz1-mum9* kernels) and (2) screen for the presence and segregation of any visible seed phenotypes. In instances where seed phenotypes are indeed visible on the ear, the parent plant is eliminated from use as a male, and its progeny is removed from the population. This allows purging of pedigrees that carry preexisting seed phenotypes. In addition to reducing build-up of mutant load, this purging minimizes propagation of non-independent seed mutations in the population and enables accurate estimation of forward seed mutation frequency each generation.

To complete development of a UniformMu line, progeny from the backcross ears are self-pollinated and scored for segregation of the *bz1-mum9* marker. About 25% of the seed on these ears will be homozygous *bz1-mum9*, and among this bronze class will be an approximate 3:1 ratio of spotted Mu-on (*MuDR* present) to non-spotted Mu-off (no *MuDR* present) phenotypes. Occasional ears will segregate 15:1 for spotted-to-stable bronze kernels if a transposition creates an unlinked duplicate of *MuDR*. Plants grown from densely spotted seed may be used as males for generating the next cycle of pedigrees, provided the parent ear shows no evidence of segregation for preexisting seed mutations.

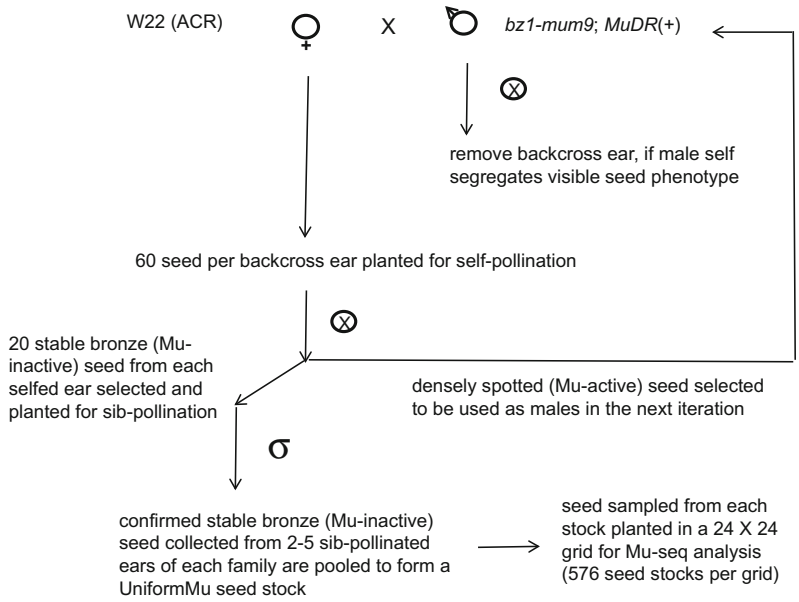

Fig. 9.2 Generation of the UniformMu population by steady-state transposon mutagenesis. The population was created by continuous backcross introgression of *MuDR* and *bz1-mum9* into the W22 (ACR) inbred. The seed stocks used in the UniformMu resource were constructed from backcross iterations four through nine

Each seed stock for public distribution is derived from stable, Mu-off, bronze kernels initially sampled from the same backcross ear. These are grown in 20-plant families and sib-pollinated to produce sustainable seed stocks. Kernels from the sib-pollinated ears are individually screened for quality (absence of mold, cracks, etc.) and complete absence of aleurone spots that indicate residual Mu-on transposase activity. In this way, seed stocks prepared for Mu-seq genotyping and distribution to users are confirmed to be Mu-off and stable throughout two generations.

While selection based on the *bz1-mum9* marker can be used for multiple generations of fast, effective screening for on-off control of Mu transposition, rare escapes can occur. Two modes may be involved. First, occasional lines exhibit complex segregation patterns consistent with a Mu-off state that derives from epigenetic silencing of *MuDR* instead of its genetic loss by Mendelian segregation (McCarty et al. 2005). Because epigenetically silenced lines have a potential to reactivate in subsequent generations, these are omitted from the UniformMu resource in favor of more stable lines (no *MuDR* rather than a silenced *MuDR*). A second mode of escape results from a change in *bz1-mum9* marker that prevents somatic reversion to *Bz1* even if *MuDR* is present. This can occur when a second mutation in the *bz1-mum9* allele preserves the bronze phenotype even when the *Mu* transposon responsible for initial mutation has excised from the *Bz1* gene. These rare events can be detected at the sequencing stage where they contribute to a low-background of non-heritable, somatic insertions. The sequences from somatic insertions are identifiable as such because the multiplex, 2D grid design used for Mu-seq leads to their detection in one, but not both axes (see below). Informatic filtering can thus remove sequences from any non-heritable somatic insertions that arise. Overall, the *bz1-mum9* marker has been a highly effective tool for achieving genetic control of the *Mutator* system.

9.2.5 Sib-pollinated Seed Stocks Maximize Preservation of Mutations

Initially, all of the new insertions generated by the steady-state protocol are in a heterozygous state. Preserving heterozygosity as much as possible is desirable for several reasons. If subsequent propagation of stable, Mu-off lines are done by repeated generations of self-pollination, about half of all heterozygous insertions will eventually be lost through segregation. Moreover, many mutations that have deleterious phenotypes when made homozygous may also be lost from the collection due to lethality, low vigor, or infertility. Finally, as a practical matter for users, segregating stocks are advantageous for obtaining immediate comparisons of mutant and wild-type individuals identifiable by gene-specific genotyping (McCarty et al. 2013b; Liu et al. 2016). If insertion lines were homozygous, researchers would need two additional generations to obtain segregating material for such comparisons; first out-crossing to W22 wild type then self-pollinating to obtain F2 progeny. At a minimum, this would entail an additional year of work for users.

In order to maximize preservation of insertions in this useful heterozygous state, seed stocks are amplified and maintained using a sib-pollination strategy. To create sustainable seed stocks, high-quality, stable (Mu-off) bronze kernels are selected from two to five, sib-pollinated ears of each family. The progeny are pooled to create a stock of 200–300 seeds that are suitable for distribution by the Maize Genetics Cooperation Stock Center. Seed stocks that become depleted can be regenerated by sib-pollination of at least five plants grown from residual seed. Once established, seed stocks are analyzed by Mu-seq to identify novel germinal *Mu* insertions (McCarty et al. 2013a). By placing Mu-seq analysis downstream of seed stock generation, we ensure that the germinal insertions detected have a high-probability of being recoverable in seed provided to users.

9.2.6 Mu-Seq Enables High-Throughput Genotyping of *Mu* Transposons

We have developed a Mu-seq protocol analogous to RNA-seq and other methods that take advantage of next generation sequencing technology. Mu-seq libraries are constructed by amplifying genomic DNA sequences that flank the highly conserved terminal-inverted-repeat (TIR) sequences located at the ends of Mu transposons. The result is a high-throughput, sequence-based method for transposon genotyping that efficiently identifies and maps germinal transposon insertions in thousands of carefully prepared seed stocks. By effectively leveraging the high read number capacity of the Illumina sequencing platform, the Mu-seq protocol enables simultaneous analysis of hundreds of UniformMu lines per library (McCarty et al. 2013a). Insertions are mapped to base-pair resolution allowing discrimination of closely spaced insertions in the genome. A typical library is constructed from a 24 × 24 grid containing 576 UniformMu lines and yields about 3100 independent germinal insertions. However, due to the accuracy, sensitivity, and flexibility of the Mu-seq approach, a variety of grid designs and pooling strategies can be employed for high-throughput genotyping applications. These include phenotype-to-genotype strategies as well as the reverse, since Mu-seq can be used for high-throughput genotyping of individuals in segregating families (Hunter et al. 2014).

9.2.7 Community Portals Provide Sustainable Access to the UniformMu Resource

MaizeGDB is the central portal for the maize genetics community (MaizeGDB.org). By providing public access to UniformMu through this well-established site, we ensure long-term sustainability of the resource. Moreover, through MaizeGDB, map locations and stock assignments of UniformMu insertions are integrated with a wealth of other genomics and genetics data that are readily searchable. MaizeGDB.org maintains an extensive database for genetic information and hosts gene browsers for sequenced inbred genomes including B73 and W22 (Springer et al. 2018). UniformMu seed stocks are deposited for distribution to the public by the Maize Genetics Cooperation Stock Center at the University of Illinois, Urbana. This center has state-of-the-art seed storage facilities designed to preserve seed viability for up to fifty years. Each insertion in the MaizeGDB.org database is linked to seed stocks that can be requested directly from the Maize Genetics Cooperation Stock Center without charge. By tradition, all user requests made to the Maize Stock Center are confidential.

9.3 Composition and Applications of the UniformMu Resource

9.3.1 Composition of the UniformMu Resource

The current UniformMu public resource contains 14,024 sustainable, seed stocks. As noted earlier, each of these stocks represents the pooled seeds of two to five sib-pollinated ears. Mu-seq analysis of these stocks has identified 74,000 independent germinal insertions. Approximately, 65% of these insertions are in or near genes. The collection includes insertions in at least 17,127 (42%) of the 39,452 maize genes in the filtered gene set (Gramene.org). The currently available UniformMu alleles also directly disrupt protein-coding sequences and probable function for at least 20% of maize genes.

Our steady-state transposon mutagenesis protocol has maintained a consistent forward mutation frequency for visible seed phenotypes of 5–7% per individual per generation in the UniformMu population (McCarty et al. 2005). While the 14,024 sequence-indexed lines have not been thoroughly phenotyped, we conservatively estimate that the collection includes >900 heritable seed phenotypes. These mutants are a rich resource for genetic dissection of seed

development (Hunter et al. 2014; McCarty 2017) and other processes. Since the resource went live on MaizeGDB in 2009, the Stock Center has fulfilled seed requests from over 1,700 users worldwide. Request rates continue to rise, and the Stock Center estimates that over 25,000 UniformMu seed packets have been distributed to researchers.

9.3.2 Applications of the UniformMu Resource

The UniformMu resource has been especially valuable for tackling complex genetic systems that potentially involve many genes. Examples span a spectrum that includes disease resistance (Yang et al. 2017a) seed development (Chen et al. 2014; Liu et al. 2013; Li et al. 2014; Shen et al. 2013; Sosso et al. 2015; Sun et al. 2015; Suzuki et al. 2006, 2008, 2015; Xiu et al. 2016; Yang et al. 2017b; Cai et al. 2017), vitamin and cofactor biosynthesis (Yang et al. 2017c), and domestication (Hufford et al. 2012; Sosso et al. 2015). The UniformMu resource provides loss-of-function alleles that are (1) readily accessible, (2) low cost, (3) available for immediate analysis, and (4) in a uniform inbred background. The lines are also useful for testing candidate genes identified by GWAS and/or QTL analyses of complex traits. Although multiple loci may control the underlying trait, and lists of candidate genes can vary in size, the roles of individual genes are ultimately most often tested by analyzing knockout mutations (Yang et al. 2017a). Since multiple candidates are typically involved, the ready availability of UniformMu insertion alleles can accelerate confirmation-testing for a subset of candidate genes.

An illustrative application of the UniformMu resource is genetic analysis of domestication and improvement phenotypes that involve many genes. Through genome-wide analysis of SNP variation in maize and teosinte, Hufford et al. (2012) identified 1,041 candidates for genes implicated in maize domestication and improvement (Table 9.1). Their analysis uncovered an additional 3,995 genes that were less directly implicated but located in genome regions that were under selection during domestication and improvement phases. As shown in Table 9.1, at least 619 of the implicated genes have coding exon-insertion (CEI) alleles in the UniformMu resource. These Mu insertions have a strong likelihood of disrupting gene function.

Table 9.1 thus presents material that will allow 619 strong, testable hypotheses to be addressed, and collectively encompasses ~20% of all candidate genes for maize domestication and improvement. As an example, Sosso et al. (2015) used two insertion alleles from UniformMu to identify *SWEET4C*, a hexose-transporter gene essential for grain-filling. The *SWEET4C* is expressed predominantly in the basal endosperm transfer cell layer of the endosperm. Loss-of-function mutations severely impact endosperm development, whereas mutant embryos can be rescued from the developing seed to produce homozygous mutant plants. Analysis of natural variation among *SWEET4C* alleles in diverse maize inbreds and teosinte accessions revealed a reduction in sequence

Table 9.1 UniformMu insertions in putative domestication genes

	Total genes	Genes with insertions	Genes with CEI alleles[a]
Domestication			
Candidate genes	468	214 (38%)	82 (18%)
Linked genes	1296	578 (45%)	243 (19%)
Improvement			
Candidates genes	573	285 (38%)	121 (21%)
Linked genes	935	404 (43%)	173 (19%)

[a]CEI (Coding Exon Insertion)

diversity upstream of *SWEET4C* in maize alleles compared to those of teosinte. Results supported SWEET4C as a probable target of selection during domestication.

While hypotheses regarding gene function are typically tested using loss-of-function mutations (e.g., insertions in protein coding exon sequences), informative dominant phenotypes have also been observed. For example, Yang et al. (2017a) initially identified a caffeoyl-CoA O-methyltransferase gene as a candidate for a QTL conferring resistance to multiple pathogens. Unexpectedly, subsequent analysis of UniformMu alleles carrying insertions in the 3′-UTR of this caffeoyl-CoA O-methyltransferase showed an enhanced, rather than a reduced disease resistance. In this case, the elevated resistance was attributed to an increase in the steady-state level of mRNA, which in turn resulted from the *Mu* insertion in the 3′-UTR. While the precise mechanism of increased mRNA stability was not addressed in this study, results clearly showed that transposon mutations with informative phenotypes were not limited to coding exon insertions alone.

9.4 Opportunities for UniformMu in the Age of Genome Editing

9.4.1 High-Throughput Phenotype to Genotype Capabilities of UniformMu

While CRISPR-based gene-editing techniques are quickly gaining favor for reverse-genetics applications (Belhaj et al. 2015), the UniformMu resource is likely to remain an important source of maize mutants. One advantage of Mu insertion alleles is that all insertions in a population can be genotyped in parallel using the high-throughput Mu-seq platform (McCarty et al. 2013a). This enables simultaneous genotyping of one or many insertions in a large population of individuals. Mu-seq is especially advantageous for forward-genetics strategies (phenotype-to-genotype) (Hunter et al. 2014). For this reason, an important goal for future development of the UniformMu resource is to include a comprehensive phenotyping of insertion lines that will facilitate the linking of genes to phenotypes using both forward- and reverse-genetics strategies.

9.4.2 Opportunities for Synergy Between UniformMu and CRISPR-Cas9

UniformMu and CRISPR-Cas9 can also interact in synergistic ways. Transposon insertions in maize genes could well be employed as universal targets for CRISPR/Cas9 based tools. One possibility is the development of a MuCRISPR line that expresses a MuCRISPR guide RNA targeting the terminal inverted repeat (TIR) sequences characteristic of Mu transposons. Such a line could provide a universal interface for a variety of Cas9 and dCas9 tools (dCas9 is a catalytically inactive Cas9 protein that binds guide RNA). Introduction of a catalytically active Cas9 into a Mu transposase system would generate deletions flanking extant Mu insertions. The result would be the generation of knockout mutants from existing lines that currently show little to no disruption of a given gene function. In addition, various dCas9 fusion proteins could be used to target genes carrying Mu insertions for activation, repression or chromatin modification (Qi et al. 2013; Piatek et al. 2015). In this way, a small set of stable transgenic maize stocks could be used to target novel modifications to any of the 17,000 genes that have UniformMu insertions in currently available lines. This strategy would also achieve a potentially large cost savings compared to that of generating a CRISPR line for each gene. Of course, as with all CRISPR/Cas9 work, appropriate controls would need to be employed to manage and identify off-target effects. In a MuCRISPR system, this would be particularly important for Cas9-generated deletions. Once created, a deletion allele of interest could typically be separated genetically from the Cas9 transgene as well as off-target mutations. If necessary, the specificity of a MuCRISPR system could be increased by targeting internal sequences of specific *Mu*

elements. Although *Mu1* insertions are prevalent in the UniformMu collection, *Mu1* is not present in the W22 (ACR) genome used as the recurrent parent. A *Mu1*-specific guide RNA would thus avoid inadvertent targeting endogenous *Mu* elements in the genome.

9.4.3 Concluding Remarks

UniformMu is fulfilling its promise as a sustainable, freely accessible, and highly useful genetic resource for maize. It will remain an important complement to emerging gene-editing approaches and may also provide novel synergisms for functional analysis of the maize genome and epigenome.

Acknowledgements This work was supported by a grant to DRM and KEK from USA National Science Foundation (IOS-1116561).

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301(5633):653–657

Barquist L, Boinett CJ, Cain AK (2013) Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biol 10(7):1161–1169

Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V (2015) Editing plant genomes with CRISPR/Cas9. Curr Opin Biotechnol 32:76–84

Bensen RJ, Johal GS, Crane VC, Tossberg JT, Schnable PS, Meeley RB, Briggs SP (1995) Cloning and characterization of the maize An1 gene. Plant Cell 7 (1):75–84

Bray RA, Brink RA (1966) Mutation and paramutation at the *R* locus in maize. Genetics 54(1):137–149

Brutnell, TP Conrad, LJ (2003). Transposon tagging using Activator (Ac) in maize. Plant Functional Genomics, Humana Press, pp 157–175

Cai M, Li S, Sun F, Sun Q, Zhao H, Ren X, Zhao Y, Tan BC, Zhang Z, Qiu F (2017) Emp10 encodes a mitochondrial PPR protein that affects the cis-splicing of nad2 intron 1 and seed development in maize. Plant J 91(1):132–144

Chen Y, Hou M, Liu L, Wu S, Shen Y, Ishiyama K, Kobayashi M, McCarty DR, Tan BC (2014) The maize DWARF1 encodes a gibberellin 3-oxidase and is dual localized to the nucleus and cytosol. Plant Physiol 166(4):2028–2039

Cresse AD, Hulbert SH, Brown WE, Lucas JR, Bennetzen JL (1995) Mu1-related transposable elements of maize preferentially insert into low copy number DNA. Genetics 140(1):315–324

Fahlgren N, Gehan MA, Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. Curr Opin Plant Biol 24:93–99

Hayes F (2003) Transposon-based strategies for microbial functional genomics and proteomics. Annu Rev Genet 37(1):3–29

Hirochika H, Guiderdoni E, An G, Hsing YI, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang Q, Pereira A, Sundaresan V (2004) Rice mutant resources for gene discovery. Plant Mol Biol 54(3):325–334

Hufford MB, Xu X, Van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44(7):808

Hunter CT, Suzuki M, Saunders J, Wu S, Tasi A, McCarty DR, Koch KE (2014) Phenotype to genotype using forward-genetic Mu-seq for identification and functional classification of maize mutants. Fron Plant Sci 4:545

Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, Kim C, Jang S, Lee S, Yang K, Nam J (2000) T-DNA insertional mutagenesis for functional genomics in rice. Plant J 22(6):561–570

Li XJ, Zhang YF, Hou M, Sun F, Shen Y, Xiu ZH, Wang X, Chen ZL, Sun SS, Small I, Tan BC (2014) Small kernel 1 encodes a pentatricopeptide repeat protein required for mitochondrial nad7 transcript editing and seed development in maize (Zea mays) and rice (Oryza sativa). Plant J 79(5):797–809

Li Y, Segal G, Wang Q, Dooner HK (2013) Gene tagging with engineered Ds elements in maize. Methods Mol Biol 1057:83–99

Lisch D, Chomet P, Freeling M (1995) Genetic characterization of the *Mutator* system in maize: behavior and regulation of *Mu* transposons in a minimal line. Genetics 139:1777–1796

Liu P, McCarty DR, Koch KE (2016) Transposon mutagenesis and analysis of mutants in UniformMu maize (Zea mays). Curr Protoc Plant Biol 451–465

Liu YJ, Xiu ZH, Meeley R, Tan BC (2013) Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. Plant Cell 25(3):868–883

May BP, Liu H, Vollbrecht E, Senior L, Rabinowicz PD, Roh D, Pan XK, Stein L, Freeling M, Alexander D, Martienssen R (2003) Maize-targeted mutagenesis: a knockout resource for maize. Proc Natl Acad Sci USA 100(20):11541–11546

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced locallesions in genomes (TILLING) for plant functional genomics. Plant Physiol 123(2): 439–442

McCarty DR, Mark Settles A, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai J, Messing J (2005) Steady-state transposon mutagenesis in inbred maize. Plant J 44(1):52–61

McCarty DR, Latshaw S, Wu S, Suzuki M, Hunter CT, Avigne WT, Koch KE (2013a) Mu-seq: sequence-based mapping and identification of transposon induced mutations. PLoS ONE 8(10):e77172

McCarty DR, Suzuki M, Hunter C, Collins J, Avigne WT, Koch KE (2013b). Genetic and molecular analyses of UniformMu transposon insertion lines. Plant Transposable Elements, Humana Press, Totowa, pp 157–166

McCarty DR (2017) 4 What Can We Learn from Maize Kernel Mutants? Maize Kernel Development, p 44

Niehaus TD, Folz J, McCarty DR, Cooper AJ, Amador DM, Fiehn O, Hanson AD (2018) Identification of a metabolic disposal route for the oncometabolite S-(2-succino) cysteine in Bacillus subtilis. J Biol Chem pp jbc–RA118

O'Malley RC, Ecker JR (2010) Linking genotype to phenotype using the Arabidopsis unimutant collection. Plant J 61(6):928–940

Piatek A, Ali Z, Baazim H, Li L, Abulfaraj A, Al-Shareef S, Aouida M, Mahfouz MM (2015) RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors. Plant Biotechnol J 13(4):578–589

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152(5):1173–1183

Raizada MN, Nan GL, Walbot V (2001) Somatic and germinal mobility of the RescueMu transposon in transgenic maize. Plant Cell 13(7):1587–1608

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL, Gao C (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31(8):686

Shen Y, Li C, McCarty DR, Meeley R, Tan BC (2013) Embryo defective12 encodes the plastid initiation factor 3 and is essential for embryogenesis in maize. Plant J 74(5):792–804

Shirley BW, Hanley S, Goodman HM (1992) Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis transparent testa mutations. Plant Cell 4 (3):333–347

Sosso D, Luo D, Li QB, Sasse J, Yang J, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS, Rogowsky PM (2015) Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. Nat Genet 47(12):1489

Springer N, W22 Genome Consortium et al (2018) The W22 genome: a foundation for maize functional genomics and transposon biology. Nat Genet https://doi.org/10.1038/s41588-018-0158-0

Sun F, Wang X, Bonnard G, Shen Y, Xiu Z, Li X, Gao D, Zhang Z, Tan BC (2015) Empty pericarp7 encodes a mitochondrial E–subgroup pentatricopeptide repeat protein that is required for ccmFN editing, mitochondrial function and seed development in maize. Plant J 84(2):283–295

Suzuki M, Mark Settles A, Tseung CW, Li QB, Latshaw S, Wu S, Porch TG, Schmelz EA, James MG, McCarty DR (2006) The maize viviparous15 locus encodes the molybdopterin synthase small subunit. Plant J 45(2):264–274

Suzuki M, Latshaw S, Sato Y, Settles AM, Koch KE, Hannah LC, Kojima M, Sakakibara H, McCarty DR (2008) The maize Viviparous8 locus, encoding a putative ALTERED MERISTEM PROGRAM1-like peptidase, regulates abscisic acid accumulation and coordinates embryo and endosperm development. Plant Physiol 146(3):1193–1206

Suzuki M, Sato Y, Wu S, Kang BH, McCarty DR (2015) Conserved functions of the MATE transporter BIG EMBRYO1 in regulation of lateral organ size and initiation rate. Plant Cell 27(8):2288–2300

Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, Conrad LJ, Kikuchi K, Kubinec TA, Hall BD, Weeks R, Unger-Wallace E, Muszynksi M, Brendel VP, Brutnell TP (2010) Genome-wide distribution of transposed dissociation elements in maize. Plant Cell 22(6):1667–1685

Walbot V, Qüesta J (2012) Using MuDR/Mu transposons in directed tagging strategies. Methods Mol Biol 1057:143–155

Williams-Carrier R, Stiffler N, Belcher S, Kroeger T, Stern DB, Monde RA, Coalter R, Barkan A (2010) Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy mutator lines of maize. Plant J 63(1):167–177

Xiu Z, Sun F, Shen Y, Zhang X, Jiang R, Bonnard G, Zhang J, Tan BC (2016) EMPTY PERICARP16 is required for mitochondrial nad2 Intron 4 cis-splicing, complex I assembly and seed development in maize. Plant J 85(4):507–519

Yang Q, He Y, Kabahuma M, Chaya T, Kelly A, Borrego E, Bian Y, El Kasmi F, Yang L, Teixeira P, Kolkman J (2017a) A gene encoding maize caffeoyl-CoA O-methyltransferase confers quantitative resistance to multiple pathogens. Nat Genet 49(9):1364

Yang YZ, Ding S, Wang HC, Sun F, Huang WL, Song S, Xu C, Tan BC (2017b) The pentatricopeptide repeat protein EMP9 is required for mitochondrial ccmB and rps4 transcript editing, mitochondrial complex biogenesis and seed development in maize. New Phytol 214(2):782–795

Yang YZ, Ding S, Wang Y, Li CL, Shen Y, Meeley R, McCarty DR, Tan BC (2017c) Small kernel2 encodes a glutaminase in vitamin B6 biosynthesis essential for maize seed development. Plant Physiol 174(2):1127–1138

Germplasm Resources for Mapping Quantitative Traits in Maize

10

Anna Glowinski and Sherry Flint-Garcia

Abstract

The expression of quantitative traits is complex, often the result of multiple genes acting in concert, and interacting with the environment. Determining the genetic control of quantitative traits can be accomplished using a number of methods to link genotype to phenotype, such as linkage-based quantitative trait locus (QTL) mapping, genome-wide association mapping (GWAS), and multi-parent mapping including nested association mapping (NAM) and multi-parent advanced generation intercrosses (MAGIC). A wide array of germplasm resources are available for mapping QTL in maize. The purpose of this chapter is to provide a brief overview of QTL mapping methods, to provide background about commonly used germplasm resources, and to discuss the strengths and weakness of each.

10.1 Introduction

Quantitative trait locus (QTL) mapping entails finding an association between a genetic marker and a measurable phenotype. Researchers work from the phenotype to the genotype, using statistical techniques to localize chromosomal regions that contain genes and/or non-coding sequences contributing to the phenotypic variation of a quantitative trait in a given population. Most traits of interest in plant breeding show quantitative inheritance, which complicates the selection process since phenotypic performances only partially reflect the genetic values of individuals. The genetic variation of a quantitative trait is controlled by the collective effects of QTL (epistasis), interactions between QTL, the environment, and QTL by environment interactions.

The goal of this chapter is to provide an overview of methods to detect genotype-to-phenotype associations for quantitative traits and to provide information about populations that are already available. This is not meant to be an exhaustive review of all traits or QTL associations as each trait group is covered in a different chapter of this book.

A. Glowinski · S. Flint-Garcia (✉)
Division of Plant Sciences, University of Missouri,
301 Curtis Hall, Columbia, MO 65211, USA
e-mail: Sherry.Flint-Garcia@ars.usda.gov

S. Flint-Garcia
USDA-ARS Plant Genetics Research Unit,
Columbia, MO 65211, USA

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_10

10.1.1 Qualitative Versus Quantitative Traits

In genetics, we can divide traits into two categories based on their effects on phenotype: qualitative and quantitative. Qualitative traits have discontinuous phenotypic variation, meaning a qualitative trait can fit into discrete categories. These are traits that can be referred to simply as "yes or no" traits, where an individual either has the trait or it does not. Usually, a single gene or a small number of genes control qualitative traits.

Conversely, quantitative traits display a continuous range of variation. Examples of quantitative traits include plant height, flowering time, and yield. These traits do not fit into discrete categories and have a continuous distribution. Generally, a larger number of genes control quantitative traits. Due to the continuous distribution of phenotypic values, quantitative genetics must employ statistical methods to link phenotypes to genotypes.

10.1.2 What Are QTL?

A QTL is a region of DNA that is associated with a particular trait, which varies in degree and which can be attributed to polygenic effects (i.e., the product of two or more genes and the environment) (Members of the Complex Trait 2003). The number of QTL which explain variation in the phenotypic trait is indicative of the genetic architecture for that trait; the more QTL, the more complex the trait.

A QTL is not a gene; at least not in the initial stages of discovery. A QTL is a large region of the genome (usually many centiMorgans and Mbp of DNA) which is linked to or contains the gene(s) that control a trait. QTL mapping is often a first step in identifying the actual genes underlying the trait because it is used to identify candidate genes in the genomic region. For genes whose function is already known, candidates can be identified based on pathways and gene expression networks. Genes of unknown function in the region can be compared to other species to identify homology-based candidates.

10.1.3 Unknown Genetic Architectures of Traits

Not all aspects of the genetic architecture of a particular trait are known. There are a variety of sources that contribute to the heritability and the genetic architecture of the trait.

10.1.3.1 Heritability

The goal of the plant breeder is to improve phenotypic values in a population by identifying and selecting superior genotypes. Because environment also affects the phenotype, there is not a perfect correspondence between phenotypic and genotypic values. To predict the outcome of selection in a collection of genotypes, a breeder must know the level of correspondence between phenotypic and genotypic values; this is known as heritability. Specifically, heritability is the percentage of the phenotypic variance that is attributable to differences among individuals in genotypic value and ranges from 0 (completely environmental) to 1 (completely genetic).

Although the heritability of a trait depends on how it is measured, in what environment(s) it is measured, and which plants are measured, different traits of maize tend to have different values of heritability. Qualitative traits, such as cob color and pericarp color, often have a value of heritability close to 1. Heritability values for quantitative traits are typically less but can vary greatly. For example, heritability can be very high for flowering (0.94; Buckler et al. 2009) and kernel protein content (0.83; Cook et al. 2012). In contrast, grain yield often has a significantly lower heritability. As a rule, traits with greater heritability can be modified more easily by selection and breeding than traits with lower heritability.

10.1.3.2 Number of Causal Loci

The term causal locus is defined as a functional genetic locus that influences and helps to explain the trait of interest. The number of causal loci contributing to a phenotype varies for different traits. Some traits are governed by many loci with smaller effects as is the case with flowering time (at least 39 QTL; Buckler et al. 2009), while

others are governed by fewer loci with larger effects (average of 4 to 5 QTL for amino acid content in grain; Deng et al. 2017). Generally, the more causal loci, the smaller the effect of each locus.

10.1.3.3 Magnitude of the Effects of Loci

Loci with larger effect sizes are more easily detected, while loci with smaller effect sizes are harder to detect. As a result of this, a large fraction of the genetic architecture of many complex traits is not well understood. Small-effect QTLs are often physically linked in a cluster or linked to large-effect QTL and fractionate during fine mapping, and there are often extensive epistatic interactions between small- and large-effect QTLs (Studer and Doebley 2011). A more complete understanding of quantitative traits will require a better understanding of the numbers, effect sizes, and genetic interactions of small-effect QTL.

10.2 Linkage-Based QTL Mapping

In linkage-based QTL mapping, QTL are mapped by identifying molecular markers that correlate with an observed trait (Veldboom et al. 1994; Grimmer et al. 2007; Zhang et al. 2009). This type of mapping depends upon recent genetic recombination between two different plant lines (as a result of a genetic cross) to identify general regions of interest.

10.2.1 Population Structure

Linkage-based QTL mapping requires the development of a mapping population, usually by crossing parents differing for the trait(s) of interest; e.g., tall x short, resistant x susceptible, high x low. The most common population structures include (1) F2 populations, (2) F2:3 populations which are created by self-pollinating the F2s to allow for replicated phenotypic trials, (3) BC1 populations where the F1 is backcrossed to the parent with the low/susceptible phenotype, (4) recombinant inbred line (RIL) populations where the F2s are self-pollinated many generations to near homozygosity in order to stabilize the genetics within each family, and (5) intermated RIL populations which allow for additional recombination prior to inbreeding. Because there is limited opportunity for recombination during the development of these population types, the linkage blocks are large and require only moderate marker density to define the recombination events.

10.2.2 Statistical Analyses

Analysis of variance (ANOVA), also known as single-marker regression, is the simplest method of linkage-based QTL mapping and was commonly used in the 1990s. The ANOVA method involves a marker regression at the marker, and provides an F statistic and associated p value for each marker. When the markers are widely spaced, the QTL may be quite far from all markers, causing low power for QTL detection.

Interval mapping makes use of a genetic map of the markers to interpolate locations between markers and, like ANOVA, assumes the presence of single QTL (Lander and Botstein 1989). Each locus is considered at one time, and the logarithm of odds ratio (LOD score) can be calculated for the comparison of two hypotheses: the presence of a QTL at a given position versus a model with no QTL at that position. A significance level is calculated by performing permutation testing (Churchill and Doerge 1994).

Composite interval mapping (CIM) can determine the location and effect size of QTL more accurately than single-QTL approaches, especially in small mapping populations where the effect of correlation between genotypes in the mapping population may be problematic. CIM is performed by using a subset of marker loci, usually identified by single-marker ANOVA, as covariates. These markers serve as proxies for unlinked QTL to increase the resolution of

interval mapping, by accounting for linked QTL and reducing the residual variation (Lynch and Walsh 1998).

10.2.3 Advantages of Linkage-Based QTL Mapping

There are many methods for linkage-based QTL mapping, and each has its advantages and disadvantages. Single-marker analysis is generally a good choice when the goal is simple detection of a QTL linked to a marker, rather than estimation of its actual position. Interval mapping offers a further increase in power of detection and more precise estimates of QTL effects and position. CIM considers the intervals between markers plus a few other well-chosen markers in each analysis, attempting to reduce or remove bias that occurs when multiple QTLs are linked to the marker/interval being considered (Lynch and Walsh 1998). One advantage all methods have in common is that linkage-based QTL mapping ***requires few genetic markers*** to ensure genome-wide coverage. In addition, depending on the population structure, the allelic classes are more balanced as compared to association analysis; see below. For example, in an F2 population, 50% of the alleles are expected to originate from each parent leading to ***higher statistical power*** per allele.

10.2.4 Disadvantages of Linkage-Based QTL Mapping

Linkage-based QTL mapping is limited to the genetic diversity present in the parents of the segregating population, leading to ***low allele richness***. Both single-marker regression and interval mapping approaches are biased when multiple QTL are linked to the maker/interval being tested. When using CIM, the biggest concern is finding suitable marker loci to serve as covariates in the analysis to help remove or reduce bias. The primary disadvantage of linkage-based approaches is the ***low mapping resolution*** due to limited recombination during population development. This low resolution can be alleviated by greatly increasing population size and/or increasing recombination through the use of advanced intercrosses.

10.3 Association Mapping

Association mapping was originally designed for the analysis of human diseases, but is now extensively used in plant genetics research as either a candidate gene association by studying single-nucleotide polymorphisms within candidate genes or as a genome-wide association study (GWAS) using anonymous molecular markers distributed across the whole genome.

Association mapping, also known as linkage disequilibrium (LD) mapping, is a method of mapping QTL that takes advantage of historic linkage disequilibrium to link phenotypes to genotypes, uncovering genetic associations (Buckler and Thornsberry 2002). It is based on the idea that polymorphisms underlying the trait that have entered a population only recently will be linked to the surrounding genetic sequence of the original evolutionary ancestor, or in other words will more often be found within a given haplotype, than outside of it.

10.3.1 Population Structure and Linkage Disequilibrium

Association mapping is generally conducted in germplasm panels consisting of pre-existing unrelated materials; i.e., no population development is required. The more diverse the germplasm is, the more rapidly LD decays within the population and the better the mapping resolution, but the more markers are required. However, if the population is too diverse, there will be a high proportion of low-frequency alleles which are either filtered out (e.g., minor allele frequencies less than 0.05) or have low statistical power. Finally, if the population has genetic structure, then the uneven distribution of alleles among the

subpopulations could lead to false positives unless population structure is accounted for in the statistical model (Yu et al. 2006). The art of assembling an association panel lies in balancing these factors.

10.3.2 Types of Association Mapping

10.3.2.1 Candidate Gene Based

As mentioned above, association mapping can be candidate gene based in which single-nucleotide polymorphisms are studied within candidate genes (Castiblanco et al. 2017). Genes associated with a phenotype of interest are selected for association mapping, and polymorphisms in only these pre-selected genes are identified and tested for association with the trait. Candidate gene-based approaches remain the most effective way of dissecting complex traits for species where sufficiently dense marker assays are not yet developed (Thavamanikumar et al. 2011), a situation that is becoming increasingly rare with the advent of next-generation sequencing-based marker systems.

10.3.2.2 Genome-Wide Association Studies (GWAS)

GWAS studies investigate the entire genome, by rapidly scanning markers across a genome to identify SNPs associated with a particular phenotype. GWAS requires corrections for population structure using PCA and/or kinship matrices in order to prevent false positives (Yu et al. 2006). Depending on the germplasm and the extent of linkage disequilibrium, GWAS generally cannot identify which polymorphisms are causal but often identifies the likely candidate gene. To date, GWAS experiments have been performed for a variety of traits in maize; see below for examples.

GWAS experiments are performed by scanning the entire genome for significant associations between a panel of SNPs and a particular phenotype. Associations must then be independently verified in order to show that they either (a) contribute to the trait of interest directly, or (b) are linked to/in linkage disequilibrium with a QTL that contributes to the trait of interest.

10.3.3 Advantages of Association Mapping

Association mapping has several advantages over linkage mapping in traditional biparental populations: (1) Currently ***existing populations are used*** rather than generating a population via a biparental cross; (2) a potentially ***large number of alleles*** per locus—compared to only two—can be surveyed simultaneously; and (3) **dramatically *increased resolution*** can be achieved (Flint-Garcia et al. 2005). Given enough statistical power and marker coverage, the low LD in maize may allow for the identification of the causative polymorphism within a candidate gene.

10.3.4 Disadvantages of Association Mapping

Association mapping requires extensive knowledge of SNP relationships within the genome, particularly in maize where LD breaks down rapidly in diverse germplasm (Flint-Garcia et al. 2003), implying that ***tens of millions of SNPs may be required*** to characterize the haplotype structure. GWAS may have ***reduced statistical power*** for detecting rare alleles because the power for detecting a QTL is determined by the frequency of alleles (Myles et al. 2009). ***False positives*** can be seen due to population structure; however, there are ways to correct for population structure (Thornsberry et al. 2001). If population structure contributes to the variation in your trait, over-correcting for population structure may lead you to many false negatives.

10.4 Nested Association Mapping

Linkage analysis and association mapping are two commonly used approaches to dissect the genetic architecture of quantitative traits (Mackay 2001;

Lander and Schork 1994). Linkage analysis and association mapping are complementary approaches: Linkage analysis often identifies large chromosome regions of interest with relatively low marker coverage, while association mapping provides high resolution with very high marker coverage (Thornsberry et al. 2001; Hirschhorn and Daly 2005). Nested association mapping (NAM) aims to create an integrated mapping population specifically designed for a full genome scan with high mapping resolution and high power for QTL with different effect sizes. The NAM strategy addresses complex trait dissection at a fundamental level through generating a common mapping resource that enables researchers to efficiently exploit genetic, genomic, and systems biology tools (Yu et al. 2008). Currently, the NAM strategy has been employed for maize (McMullen et al. 2009) and other species (Fragoso et al. 2017; Song et al. 2017; Bajgain et al. 2016), with NAM populations under development for many other species.

10.4.1 Advantages of NAM

NAM takes advantage of both historic and recent recombination events in order to achieve ***low marker density requirements***, ***high allele richness***, ***high mapping resolution***, and ***high statistical power***, with none of the disadvantages of either linkage analysis or association mapping. This allows for the discovery of QTL with greater precision and accuracy. Parental alleles are shuffled over several generations through segregation and genetic recombination providing new combinations of alleles for study. NAM populations also have an added benefit, in that they can function as an archive for genetic diversity.

10.4.2 Disadvantages of NAM

Unless a NAM population already exists as in the case of maize, a new ***population must be generated*** which utilizes both time and resources. Challenges include ensuring that the pedigree of each cross is maintained while advancing to the next generation, that the founders are diverse enough to carry different alleles for important characteristics, and that near-complete homozygosity is reached in the final population. While NAM captures thousands of recombination events, recombination and segregation distortion vary among different families which can limit the precision of genetic dissection of quantitative traits (McMullen et al. 2009; Ladejobi et al. 2016).

10.5 Available QTL Mapping Populations and Germplasm Resources

10.5.1 Intermated B73/Mo17 (IBM)

Hundreds of linkage-based QTL populations have been created over the past 30 years; each is focused on a specific trait(s) but rarely made available to the public. However, the maize community in the public sector has championed the use of the intermated B73 × Mo17 (IBM) population as a central linkage mapping resource because of the historical value of the two parents and the value of merging the genetic map (IBM) to the maize genome (B73) (Coe et al. 2002).

The IBM population was the first widely used QTL population derived from additional generations of intermating prior to inbreeding (Lee et al. 2002). The IBM is comprised of approximately 300 RILs, 94 of which are referred to as "the core set." The RILs were derived from the single-cross hybrid of inbreds B73 (female) and Mo17 (Lee et al. 2002). A single F1 plant was self-pollinated to produce the F2 generation. In the F2, plants were used once, as male or female, in a cross with another plant so that 250 pairs of plants were mated. A single kernel was taken from each ear and bulked with the seed of the other ears to form the Syn1 generation. This procedure was repeated for four additional generations to produce the Syn5 generation. The increased opportunity for recombination in IBM has resulted in an almost four-fold increase in the

genetic map distance compared with conventional non-intermated RIL populations, allowing for more precise definition of QTLs. IBM has been widely used for developing genetic markers and anchoring them to the genetic map as well as the physical map (Coe et al. 2002) and for studying the genetic architecture of numerous traits (e.g., Eichten et al. 2011; Ordas et al. 2009; Rodríguez et al. 2008; Zhang et al. 2010a; Baxter et al. 2013; Balint-Kurti et al. 2007; Dubois et al. 2010; Hazen et al. 2003).

10.5.2 Association Panels

Describing the commonly used association panels is a somewhat difficult task. They do not require additional population development as they are typically collections of materials previously created by multiple groups. Because of this, association panels are extremely easy to modify by merging panels together, dropping various groups of germplasm from a panel based on phenology (e.g., adaptation to temperate or tropical environments) and/or germplasm availability (e.g., not all germplasm is publicly available), and customizing panels for specific traits by adding lines chosen for extremes in the trait. Studies often report phenotypes on multiple panels for the same trait(s), and the results are compared in the context of allele frequencies and population structure. The following is a short list of the most commonly used association panels.

10.5.2.1 Maize 282 Association Panel

The maize 282 association panel was one the earliest association panels in maize and consists of breeding lines assembled by Major Goodman (Flint-Garcia et al. 2005). The very first maize association panel by Thornsberry et al. (2001) consisted of 102 inbred lines, but it was quickly realized that this small of a panel had insufficient power to detect QTL; hence, it was increased to 302 inbred lines (Flint-Garcia et al. 2005) based on pedigree information (e.g., Gerdes et al. 1993) and prior to the availability of SNPs to characterize germplasm relationships. After genotyping, a number of isolines (highly related lines derived from backcrossing with B73, for example) were identified and removed from the panel yielding the current 282 association panel. The 282 panel, also known as the Goodman–Buckler panel, represents a sample of the diversity present in the public sector including current breeding lines (at the time of development) as well as historically important lines from both temperate and tropical programs. This association panel has been used for a variety of association studies since its creation (e.g., Krill et al. 2010; Hung et al. 2012; Cook et al. 2012; Hu et al. 2018; Diepenbrock et al. 2017; Hu et al. 2017; Olukolu et al. 2016; Zhang et al. 2010b; Butron et al. 2010; Harjes et al. 2008; Benke et al. 2015; Samayoa et al. 2015; Olukolu et al. 2013).

10.5.2.2 Ames Association Panel

The USDA North Central Regional Plant Introduction Station (NCRPIS) in Ames, Iowa, maintains over 3000 maize inbreds from around the world. The Ames panel was created by choosing over 2500 inbred lines from the NCRPIS inbred collection based only on sufficient seed availability and a minimum of five generations of self-pollination to ensure an inbred nature, and represents nearly a century of maize breeding efforts. The panel has been genotypically characterized by genotyping-by-sequencing (GBS; Elshire et al. 2011) in order to assist with curatorial management of germplasm collections and to evaluate diversity within breeding programs (Romay et al. 2013) and for use in association mapping (Lu et al. 2015; Xue et al. 2016; Peiffer et al. 2014; Zila et al. 2013). Because the population is so large and genetically diverse, subsets of lines from the Ames association panel have been used successfully to characterize many different traits (Pace et al. 2015).

10.5.2.3 Wisconsin Diversity Panel

A subset of the Ames panel with a reduced phenology for adaptation to the northern corn belt was chosen and is known as the Wisconsin Diversity Panel (WiDiv). This panel contains 627 lines selected based on flowering in the target

environment of Wisconsin, agronomic suitability, uniformity, and seed supply (Hansey et al. 2011).

Many of the WiDiv panel lines trace back to eight open-pollinated populations including Iowa Stiff Stalk, Minnesota No. 13, Reid Yellow Dent, Lancaster Surecrop, Golden Glow, Funk Yellow Dent, Pride of Saline, and Krug among others. Having multiple genotypes derived from the same open-pollinated population helps to maintain a balance of allele frequencies which results in increased statistical power. The WiDiv has been used in a number of high throughput image analysis projects to investigate stalk, tassel, ear, and kernel morphology traits (Miller et al. 2017; Gustin et al. 2013; Heckwolf et al. 2015; Muttoni et al. 2012), and for GWAS studies of juvenile-to-adult vegetative and vegetative-to-reproductive developmental transitions (Hirsch et al. 2014b) and tassel traits (Gage et al. 2018).

10.5.2.4 CIMMYT Association Panels

The International Center for Maize and Wheat Improvement (CIMMYT) has a global mandate for improving the productivity and sustainability of maize and wheat in developing countries (Hoisington et al. 1999). The CIMMYT maize germplasm bank contains over 28,000 seed samples, including inbred lines, breeding populations, landraces, and wild relatives. To leverage this germplasm resource, CIMMYT has developed a number of GWAS panels, primarily to study grain carotenoid content and drought tolerance. The carotenoid research was conducted on two panels: a set of 245 diverse maize inbred lines predominantly derived from tropical and subtropical adapted maize germplasm (Yan et al. 2010) and the carotenoid association mapping (CAM) panel consisting of 380 primarily diverse tropical and subtropical lines assembled by the HarvestPlus-funded program at CIMMYT (Suwarno et al. 2015).

The CIMMYT drought panel of 350 inbred lines was used to test candidates in abscisic acid (ABA) in response to drought (Setter et al. 2011) and to conduct GWAS for agronomic trait (Xue et al. 2013) and metabolic (Zhang et al. 2016) responses to drought as compared to well-watered conditions. In addition, the drought-tolerant maize for Africa (DTMA; ~250–300 lines) and improved maize for African soils (IMAS; ~400 lines) panels were combined to identify a major QTL for resistance to tar spot complex (Mahuku et al. 2016) and maize lethal necrosis disease (Gowda et al. 2015). Finally, another collection of 940 African lines was genotyped (Semagn et al. 2012) and evaluated disease resistance including Fusarium ear rot (Chen et al. 2016).

10.5.2.5 Chinese Association Panels

The first Chinese association panel was composed of 155 diverse temperate-adapted maize inbred lines from China (Yang et al. 2010) and was later referred to as the Chinese association mapping (CAM155) panel in subsequent publications by the lead authors (Li et al. 2011). This panel was used in a GWAS study of kernel carotenoids (Yan et al. 2010), before being merged with other germplasm to form additional panels such as the AM508 (see below).

A broader global diverse line panel of 527/513 inbred lines representative of tropical, subtropical, and temperate germplasm was collected to construct a larger association panel (Yang et al. 2011). This collection includes 527 lines from the GEM project, CIMMYT maize breeding programs, elite parents of commercial hybrids widely used in China, lines derived from Chinese landraces, and high-oil and high-provitamin A lines; 513 lines were genotyped with the Illumina MaizeSNP50 array and used for GWAS of kernel α-tocopherol content (Li et al. 2012), maize rough dwarf mosaic virus (Chen et al. 2015), and a large number of plant, ear, kernel, and yield-related traits (Yang et al. 2014), among many other traits.

More recently, the 527-/513-inbred line panel above was reduced to a 508-inbred line panel known as AM508 which was first described in an investigation of kernel oil content (Li et al. 2013). A commonly used subset of the AM508 is a set of 368 lines which was subjected to RNA-seq (Fu et al. 2013). This 368-line panel was used to investigate many traits and phenomena such as expression QTL (eQTL),

regulatory networks, non-coding sequences, and metabolites in the developing kernel (Fu et al. 2013; Wen et al. 2014; Liu et al. 2017).

10.5.2.6 European Association Panels

A widely used panel in Europe consisted of 375 inbred lines representative of American, European, and tropical maize (Camus-Kulandaivelu et al. 2006) which included the original 102-inbred subset of the 282 (Thornsberry et al. 2001) and a unique set of 153 inbreds derived from self-pollinating European landraces. This panel was used to study epistatic interactions in *Opaque2* for kernel traits (Manicacci et al. 2009), flowering time (Durand et al. 2012; Camus-Kulandaivelu et al. 2006), and phenology and plant architecture traits (Bouchet et al. 2016).

A set of 289 diverse dent inbred lines from the Americas, Europe, and China has been assembled to investigate genomic and metabolic prediction of heterosis (Riedelsheimer et al. 2012a) as well as GWAS for leaf metabolites and biomass-related traits (Riedelsheimer et al. 2012b).

Another recent association panel, of sorts, is comprised of two separate panels of 306 dent and 292 flint maize inbred lines based on collections of Spanish, French, and German breeders from the Cornfed Project. These are often evaluated as hybrids with the opposite heterotic group (i.e., flint panel crossed with dent tester and vice versa). This panel has been investigated for cold tolerance (Revilla et al. 2016).

10.5.2.7 Other Association Panels

Private industry has also used association analysis using their propriety germplasm, though few GWAS studies have been published by industry. Of those published, Belo et al. (2008) used 553 historically important and current elite maize inbred lines from Pioneer Hi-Bred to conduct GWAS for fatty acid content in kernels; 1,487 inbred lines from Limagrain representing elite European and North American germplasm were used to investigate northern corn leaf blight (Van Inghelandt et al. 2012); and Dow AgroSciences used 300 inbreds, including 215 Dow proprietary lines of North and South American origin, to validate QTL for gray leaf spot (Mammadov et al. 2015).

Additional trait- and geography-specific panels have been assembled. For example, the 300-inbred line panel of Warburton et al. (2013) was used to investigate resistance *Aspergillus flavus*, aflatoxin accumulation, and drought (Warburton et al. 2013; Farfan et al. 2015). A subset of 287 these 300 lines has been used to map resistance to corn earworm and associated metabolic pathways (Warburton et al. 2018). A Brazilian panel of 183 lines was assembled to conduct GWAS for Fusarium ear rot resistance (Coan et al. 2018), but is being expanded to 335 (M. Warburton, personal communication). A set of 240 Indian and CIMMYT lines were analyzed for associations with yield and yield component traits under drought conditions (Thirunavukkarasu et al. 2014).

10.5.3 US NAM

The US NAM consists of 5000 RILs derived by crossing 25 diverse maize lines to B73 (McMullen et al. 2009). The 25 diverse inbred lines were chosen as parents to maximize diversity encompassed in the 282 association panel (Flint-Garcia et al. 2005) and preserve historic linkage disequilibrium (Yu et al. 2008). Each parental line was crossed to B73, the inbred chosen for the reference genome. The F1 plants were then self-pollinated by single seed descent for six generations to create a total of 200 homozygous RILs per family, for a total of 5000 RILs which were originally genotyped with 1536 SNPs (McMullen et al. 2009) and subsequently by GBS. Benefits of using the US NAM population for QTL mapping include broader genetic diversity, higher mapping resolution than individual biparental populations, and an increase in statistical power because allele frequencies are balanced within each family. While the US NAM population only taps the diversity of 25 founder lines, it is large enough to address questions regarding magnitudes of QTL effects, heterosis, and the mapping of numerous genes controlling various traits (e.g., Buckler et al. 2009; Poland

et al. 2011; Cook et al. 2012; Hirsch et al. 2014a; Handrick et al. 2016; Kump et al. 2011; Tian et al. 2011; Benson et al. 2015; Brown et al. 2011).

10.5.4 European NAM

The European NAM population was created by creating two half-sib panels of 11 and 13 half-sib families, one for European Dent and one for European Flint maize, respectively (Bauer et al. 2013). Each of the two panels consists of a common parent crossed to founder lines that represent important and diverse breeding lines of the European maize germplasm. In the Dent panel, a central line (F353 from France) was crossed with ten Dent founder lines. In the Flint panel, the central line (UH007 from Germany) was crossed with 11 Flint founder lines. In addition, each of the common parents was crossed with B73, and the reciprocal populations F353xUH007 and UH007xF353 were generated. These additional populations were made to connect the two panels to each other and with the US NAM population. All progenies are homozygous doubled haploid lines obtained from F1 plants. The resulting 24 doubled haploid populations each consist of 35–129 lines, for a total of 2,267 doubled haploid lines, and have been genotyped with the Illumina MaizeSNP50 array (Ganal et al. 2011). The European NAM population has been used to study recombination rate (Bauer et al. 2013) and genomic prediction of yield (Lehermeier et al. 2014).

10.5.5 Chinese NAM

The Chinese NAM (CN-NAM) population was developed by crossing 11 diverse inbred lines representing the heterotic groups used in Chinese maize breeding with the common parent "HZS" which has wide adaptability and good combining ability. The F2s were self-pollinated to create 1971 RILs which were genotyped by GBS (Li et al. 2015). The CN-NAM population has been used to dissect drought tolerance (Li et al. 2016a), inflorescence size (Wu et al. 2016), and flowering time (Li et al. 2015, 2016b).

10.5.6 Multi-parent Advanced Generation Intercrosses (MAGIC) and Other Multi-parent Populations

Multi-parent advanced generation intercross (MAGIC) populations have now been developed for a variety of species including maize, rice, wheat, and *Arabidopsis* (Dell'Acqua et al. 2015; Bandillo et al. 2013; Huang et al. 2012; Kover et al. 2009). The maize MAGIC population contains 1636 RILs derived from eight genetically diverse founder lines that were crossed in a funnel breeding design (Dell'Acqua et al. 2015). RILs were produced by pooling two-way, four-way, and eight-way hybrids in 35 independent breeding funnels (subfamilies). Each funnel was advanced by single seed descent to the F6 generation. This MAGIC population has been used to investigate flowering time, plant and ear height and grain size (Dell'Acqua et al. 2015).

The mapping power and resolution of MAGIC maize are strengthened by high minor allele frequencies and a rapid decay of linkage disequilibrium. Similar to the US NAM population, MAGIC maize has broader genetic diversity, higher resolutions than biparental populations, and a reduction of problems associated with the frequency of rare alleles (Holland 2015). These benefits make MAGIC maize a useful population for QTL mapping in maize.

A more recent multi-parent mapping method is called random-open-parent association mapping (ROAM), where RIL populations are derived from a number of inbred lines crossed in combinations without an a priori requirement to interconnect across populations (Xiao et al. 2016; Pan et al. 2016). In concept, the ROAM method could be used when merging multiple NAM panels together, such as the US, CN, and European NAM populations, where different common hub parents are used and where a variable number of populations are derived from each hub. An advantage of ROAM is that additional

families can be developed and added, as there is no a priori design to the larger ROAM population. A possible disadvantage is that there may be lower power for some alleles due to unbalanced allele frequencies among the parents (Xiao et al. 2017).

10.6 Conclusions

The goal of this chapter was to provide an overview of methods for genotype-to-phenotype associations and introduce some of the mapping resources that are available for study. Detailed reviews of QTL and GWAS analyses for insect resistance, fungal diseases, cold tolerance, root system architecture traits, nitrogen use efficiency, and kernel oil content can be found in other chapters of this book.

References

Bajgain P, Rouse MN, Tsilo TJ, Macharia GK, Bhavani S, Jin Y, Anderson JA (2016) Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. PLOS ONE 11(5):e0155760. https://doi.org/10.1371/journal.pone.0155760

Balint-Kurti PJ, Zwonitzer JC, Wisser RJ, Carson ML, Oropeza-Rosas MA, Holland JB, Szalma SJ (2007) Precise mapping of quantitative trait loci for resistance to southern leaf blight, caused by Cochliobolus heterostrophus race O, and flowering time using advanced intercross maize lines. Genetics 176:645–657

Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C-W, McCouch S, Thomson M, Mauleon R, Singh RK, Gregorio G, Redoña E, Leung H (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6(1):11. https://doi.org/10.1186/1939-8433-6-11

Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L, Meyer N, Ranc N, Rincent R, Schipprack W, Altmann T, Flament P, Melchinger AE, Menz M, Moreno-González J, Ouzunova M, Revilla P, Charcosset A, Martin OC, Schön C-C (2013) Intraspecific variation of recombination rate in maize. Genome Biol 14(9):R103. https://doi.org/10.1186/gb-2013-14-9-r103

Baxter I, Gustin JL, Settles M, Hoekenga OA (2013) Ionomic characterization of maize kernels in the intermated B73 × Mo17 population. Crop Sci 3:208–220

Belo A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, Tingey S, Rafalski A (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Mol Genet Genomics 279(1):1–10

Benke A, Urbany C, Stich B (2015) Genome-wide association mapping of iron homeostasis in the maize association population. BMC Genet 16(1):1. https://doi.org/10.1186/s12863-014-0153-0

Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ (2015) Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. PLOS Genetics 11(3):e1005045. https://doi.org/10.1371/journal.pgen.1005045

Bouchet S, Bertin P, Presterl T, Jamin P, Coubriche D, Gouesnard B, Laborde J, Charcosset A (2016) Association mapping for phenology and plant architecture in maize shows higher power for developmental traits compared with growth influenced traits. Heredity 118:249. https://doi.org/10.1038/hdy.2016.88. https://www.nature.com/articles/hdy201688#supplementary-information

Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, Holland JB, Flint-Garcia S, McMullen MD, Buckler ES, Rocheford TR (2011) Distinct genetic architectures for male and female inflorescence traits of maize. PLoS Genet 7(11):e1002383. https://doi.org/10.1371/journal.pgen.1002383

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC (2009) The genetic architecture of maize flowering time. Science 325(5941):714–718

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5(2):107–111. https://doi.org/10.1016/S1369-5266(02)00238-8

Butron A, Chen YC, Rottinghaus GE, McMullen MD (2010) Genetic variation at bx1 controls DIMBOA content in maize. Theor Appl Genet 120:721–734

Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172(4):2449–2463

Castiblanco V, Marulanda JJ, Würschum T, Miedaner T (2017) Candidate gene based association mapping in Fusarium culmorum for field quantitative pathogenicity and mycotoxin production in wheat. BMC Genet 18(1):49. https://doi.org/10.1186/s12863-017-0511-9

Chen G, Wang X, Hao J, Yan J, Ding J (2015) Genome-wide association implicates candidate genes conferring resistance to maize rough dwarf disease in maize. PLOS ONE 10(11):e0142001. https://doi.org/10.1371/journal.pone.0142001

Chen J, Shrestha R, Ding J, Zheng H, Mu C, Wu J, Mahuku G (2016) Genome-wide association study and QTL mapping reveal genomic loci associated with *Fusarium* ear rot resistance in tropical maize

germplasm. G3 Genes Genomes Genet 6(12):3803–3815. https://doi.org/10.1534/g3.116.034561

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138: 963–971

Coan MMD, Senhorinho HJC, Pinto RJB, Scapim CA, Tessmann DJ, Williams WP, Warburton ML (2018) Genome-wide association study of resistance to ear rot by Fusarium verticillioides in a tropical field maize and popcorn core collection. Crop Sci 58(2):564–578. https://doi.org/10.2135/cropsci2017.05.0322

Coe E, Cone K, McMullen M, Chen S-S, Davis G, Gardiner J, Liscum E, Polacco M, Paterson A, Sanchez-Villeda H, Soderlund C, Wing R (2002) Access to the maize genome: an integrated physical and genetic map. Plant Physiol 128(1):9–12. https://doi.org/10.1104/pp.010953

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158(2):824–834. https://doi.org/10.1104/pp.111.185033

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J, Frascaroli E, Churchill GA, Inzé D, Morgante M, Pè ME (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol 16(1):167. https://doi.org/10.1186/s13059-015-0716-z

Deng M, Li D, Luo J, Xiao Y, Liu H, Pan Q, Zhang X, Jin M, Zhao M, Yan J (2017) The genetic architecture of amino acids dissection by association and linkage analysis in maize. Plant Biotechnol J 15(10):1250–1263. https://doi.org/10.1111/pbi.12712

Diepenbrock CH, Kandianis CB, Lipka AE, Magallanes-Lundback M, Vaillancourt B, Góngora-Castillo E, Wallace JG, Cepela J, Mesberg A, Bradbury PJ, Ilut DC, Mateos-Hernandez M, Hamilton J, Owens BF, Tiede T, Buckler ES, Rocheford T, Buell CR, Gore MA, DellaPenna D (2017) Novel loci underlie natural variation in vitamin E levels in maize grain. Plant Cell 29(10):2374–2392. https://doi.org/10.1105/tpc.17.00475

Dubois PG, Olsefski GT, Flint-Garcia S, Setter TL, Hoekenga OA, Brutnell TP (2010) Physiological and genetic characterization of end-of-day far-red light response in maize seedlings. Plant Physiol 154 (1):173–186. https://doi.org/10.1104/pp.110.159830

Durand E, Bouchet S, Bertin P, Ressayre A, Jamin P, Charcosset A, Dillmann C, Tenaillon MI (2012) Flowering time in maize: linkage and epistasis at a major effect locus. Genetics 190:1547–1562. https://doi.org/10.1534/genetics.111.136903

Eichten SR, Foerster JM, de Leon N, Kai Y, Yeh C-T, Liu S, Jeddeloh JA, Schnable PS, Kaeppler SM, Springer NM (2011) B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. Plant Physiol 156(4):1679–1690. https://doi.org/10.1104/pp.111.174748

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):e19379. https://doi.org/10.1371/journal.pone.0019379

Farfan IDB, De La Fuente GN, Murray SC, Isakeit T, Huang P-C, Warburton M, Williams P, Windham GL, Kolomiets M (2015) Genome wide association study for drought, aflatoxin resistance, and important agronomic traits of maize hybrids in the sub-tropics. PLOS ONE 10(2):e0117737. https://doi.org/10.1371/journal.pone.0117737

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44(6):1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Fragoso CA, Moreno M, Wang Z, Heffelfinger C, Arbelaez LJ, Aguirre JA, Franco N, Romero LE, Labadie K, Zhao H, Dellaporta SL, Lorieux M (2017) Genetic architecture of a rice nested association mapping population. G3 Genes Genomes Genet 7(6): 1913–1926. https://doi.org/10.1534/g3.117.041608

Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z, Wang B, Zhai L, Dai C, Xu J, Wang W, Li X, Zheng J, Chen L, Luo L, Liu J, Qian X, Yan J, Wang J, Wang G (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. Nat Commun 4:2832. https://doi.org/10.1038/ncomms3832. https://www.nature.com/articles/ncomms3832#supplementary-information

Gage JL, White MR, Edwards J, Kaeppler S, de Leon N (2018) Selection signatures underlying dramatic male inflorescence transformation during modern hybrid maize breeding. bioRxiv. https://doi.org/10.1101/284109

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner EM, Hansen M, Joets J, Le Paslier MC, McMullen MD, Montalent P, Rose M, Schon CC, Sun Q, Walter H, Martin OC, Falque M (2011) A large maize (Zea mays L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6. https://doi.org/10.1371/journal.pone.0028334

Gerdes JT, Behr CF, Coors JG, Tracy WF (1993) Compilation of North American maize breeding germplasm. CSSA, Madison, WI

Gowda M, Das B, Makumbi D, Babu R, Semagn K, Mahuku G, Olsen MS, Bright JM, Beyene Y, Prasanna BM (2015) Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor Appl Genet 128(10):1957–1968. https://doi.org/10.1007/s00122-015-2559-0

Grimmer MK, Trybush S, Hanley S, Francis SA, Karp A, Asher MJC (2007) An anchored linkage map for sugar

beet based on AFLP, SNP and RAPD markers and QTL mapping of a new source of resistance to Beet necrotic yellow vein virus. Theor Appl Genet 114 (7):1151–1160. https://doi.org/10.1007/s00122-007-0507-3

Gustin JL, Jackson S, Williams C, Patel A, Armstrong P, Peter GF, Settles AM (2013) Analysis of maize (Zea mays) kernel density and volume using microcomputed tomography and single-kernel near-infrared spectroscopy. J Agric Food Chem 61(46):10872–10880. https://doi.org/10.1021/jf403790v

Handrick V, Robert CAM, Ahern KR, Zhou S, Machado RAR, Maag D, Glauser G, Fernandez-Penny FE, Chandran JN, Rodgers-Melnik E, Schneider B, Buckler ES, Boland W, Gershenzon J, Jander G, Erb M, Köllner TG (2016) Biosynthesis of 8-*O*-methylated benzoxazinoid defense compounds in maize. Plant Cell 28(7):1682–1700. https://doi.org/10.1105/tpc.16.00065

Hansey CN, Johnson JM, Sekhon RS, Kaeppler SM, De Leon N (2011) Genetic diversity of a maize association population with restricted phenology. Crop Sci 51. https://doi.org/10.2135/cropsci2010.03.0178

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319(5861): 330–333

Hazen SP, Hawley RM, Davis GL, Henrissat B, Walton JD (2003) Quantitative trait loci and comparative genomics of cereal cell wall composition. Plant Physiol 132:263–271

Heckwolf S, Heckwolf M, Kaeppler SM, de Leon N, Spalding EP (2015) Image analysis of anatomical traits in stalk transections of maize and other grasses. Plant Methods 11(1):26. https://doi.org/10.1186/s13007-015-0070-x

Hirsch CN, Flint-Garcia SA, Beissinger TM, Eichten SR, Deshpande S, Barry K, McMullen MD, Holland JB, Buckler ES, Springer N, Buell CR, de Leon N, Kaeppler SM (2014a) Insights into the effects of long-term artificial selection on seed size in maize. Genetics 198(1):409–421. https://doi.org/10.1534/genetics.114.167155

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N, Kaeppler SM, Buell CR (2014b) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26(1):121–135. https://doi.org/10.1105/tpc.113.119982

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95. https://doi.org/10.1038/nrg1521

Hoisington D, Khairallah M, Reeves T, Ribaut J-M, Skovmand B, Taba S, Warburton M (1999) Plant genetic resources: what can they contribute toward increased crop productivity? Proc Natl Acad Sci 96 (11):5937–5943. https://doi.org/10.1073/pnas.96.11.5937

Holland JB (2015) MAGIC maize: a new resource for plant genetics. Genome Biol 16(1):163. https://doi.org/10.1186/s13059-015-0713-2

Hu G, Li Z, Lu Y, Li C, Gong S, Yan S, Li G, Wang M, Ren H, Guan H, Zhang Z, Qin D, Chai M, Yu J, Li Y, Yang D, Wang T, Zhang Z (2017) Genome-wide association study Identified multiple genetic loci on chilling resistance during germination in maize. Sci Rep 7(1):10840. https://doi.org/10.1038/s41598-017-11318-6

Hu Y, Ren J, Peng Z, Umana AA, Le H, Danilova T, Fu J, Wang H, Robertson A, Hulbert SH, White FF, Liu S (2018) Analysis of extreme phenotype bulk copy number variation (XP-CNV) identified the association of rp1 with resistance to Goss's wilt of maize. Front Plant Sci 9(110). https://doi.org/10.3389/fpls.2018.00110

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol J 10. https://doi.org/10.1111/j.1467-7652.2012.00702.x

Hung H-Y, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF, Holland JB (2012) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc Natl Acad Sci 109(28):E1913–E1921. https://doi.org/10.1073/pnas.1203189109

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLOS Genet 5(7):e1000551. https://doi.org/10.1371/journal.pgen.1000551

Krill AM, Kirst M, Kochian LV, Buckler ES, Hoekenga OA (2010) Association and linkage analysis of aluminum tolerance genes in maize. PLOS ONE 5(4):e9958. https://doi.org/10.1371/journal.pone.0009958

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43. https://doi.org/10.1038/ng.747

Ladejobi O, Elderfield J, Gardner KA, Gaynor RC, Hickey J, Hibberd JM, Mackay IJ, Bentley AR (2016) Maximizing the potential of multi-parental crop populations. Appl Transl Genomics 11(Supplement C):9–17. https://doi.org/10.1016/j.atg.2016.10.002

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265(5181):2037–2048

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 × Mo17 (IBM) population. Plant Mol Biol 48(5):453–461. https://doi.org/10.1023/a:1014893521186

Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger AE, Menz M, Meyer N, Moreau L, Moreno-González J, Ouzunova M, Pausch H, Ranc N, Schipprack W, Schönleben M, Walter H, Charcosset A, Schön C-C (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. Genetics 198(1):3–16. https://doi.org/10.1534/genetics.114.161943

Li C, Li Y, Bradbury PJ, Wu X, Shi Y, Song Y, Zhang D, Rodgers-Melnick E, Buckler ES, Zhang Z, Li Y, Wang T (2015) Construction of high-quality recombination maps with low-coverage genomic sequencing for joint linkage analysis in maize. BMC Biol 13 (1):78. https://doi.org/10.1186/s12915-015-0187-4

Li C, Sun B, Li Y, Liu C, Wu X, Zhang D, Shi Y, Song Y, Buckler ES, Zhang Z, Wang T, Li Y (2016a) Numerous genetic loci identified for drought tolerance in the maize nested association mapping populations. BMC Genomics 17(1):894. https://doi.org/10.1186/s12864-016-3170-8

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y, Wang G, Li J, Yan J (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45(1):43–50. http://www.nature.com/ng/journal/v45/n1/abs/ng.2484.html#supplementary-information

Li L, Li H, Li Q, Yang X, Zheng D, Warburton M, Chai Y, Zhang P, Guo Y, Yan J, Li J (2011) An 11-bp Insertion in Zea mays fatb reduces the palmitic acid content of fatty acids in maize grain. PLoS ONE 6(9): e24699. https://doi.org/10.1371/journal.pone.0024699

Li Q, Yang X, Xu S, Cai Y, Zhang D, Han Y, Li L, Zhang Z, Gao S, Li J, Yan J (2012) Genome-wide association studies identified three independent polymorphisms associated with α-tocopherol content in maize kernels. PLOS ONE 7(5):e36807. https://doi.org/10.1371/journal.pone.0036807

Li YX, Li C, Bradbury PJ, Liu X, Lu F, Romay CM, Glaubitz JC, Wu X, Peng B, Shi Y, Song Y, Zhang D, Buckler ES, Zhang Z, Li Y, Wang T (2016b) Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. Plant J 86(5):391–402. https://doi.org/10.1111/tpj.13174

Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, Wang X, Jin M, Li W, Zhang Q, Yan J (2017) Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. Mol Plant 10(3):414–426. https://doi.org/10.1016/j.molp.2016.06.016

Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, Li Y, Li Y, Semagn K, Zhang X, Hernandez AG, Mikel MA, Soifer I, Barad O, Buckler ES (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. Nat Commun 6:6914. https://doi.org/10.1038/ncomms7914. https://www.nature.com/articles/ncomms7914#supplementary-information

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer

Mackay TFC (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35(1):303–339. https://doi.org/10.1146/annurev.genet.35.102401.090633

Mahuku G, Chen J, Shrestha R, Narro LA, Guerrero KVO, Arcos AL, Xu Y (2016) Combined linkage and association mapping identifies a major QTL (qRtsc8-1), conferring tar spot complex resistance in maize. Theor Appl Genet 129(6):1217–1229. https://doi.org/10.1007/s00122-016-2698-y

Mammadov J, Sun X, Gao Y, Ochsenfeld C, Bakker E, Ren R, Flora J, Wang X, Kumpatla S, Meyer D, Thompson S (2015) Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (Zea mays L.). BMC Genomics 16(1):916. https://doi.org/10.1186/s12864-015-2171-3

Manicacci D, Camus-Kulandaivelu L, Fourmann M, Arar C, Barrault S, Rousselet A, Feminias N, Consoli L, Francès L, Méchin V, Murigneux A, Prioul J-L, Charcosset A, Damerval C (2009) Epistatic Interactions between *Opaque2* transcriptional activator and its target gene *CyPPDK1* control kernel trait variation in maize. Plant Physiol 150(1):506–520. https://doi.org/10.1104/pp.108.131888

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q (2009) Genetic properties of the maize nested association mapping population. Science 325. https://doi.org/10.1126/science.1174320

Members of the Complex Trait C (2003) The nature and identification of quantitative trait loci: a community's view. Nat Rev Genet 4:911. https://doi.org/10.1038/nrg1206. https://www.nature.com/articles/nrg1206#supplementary-information

Miller ND, Haase NJ, Lee J, Kaeppler SM, Leon N, Spalding EP (2017) A robust, high-throughput method for computing maize ear, cob, and kernel attributes automatically from images. Plant J 89(1):169–178. https://doi.org/10.1111/tpj.13320

Muttoni G, Johnson JM, Santoro N, Rhiner CJ, von Mogel KJH, Kaeppler SM, de Leon N (2012) A high-throughput core sampling device for the evaluation of maize stalk composition. Biotechnol Biofuels 5(1):27. https://doi.org/10.1186/1754-6834-5-27

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21(8):2194–2202. https://doi.org/10.1105/tpc.109.068437

Olukolu BA, Bian Y, De Vries B, Tracy WF, Wisser RJ, Holland JB, Balint-Kurti PJ (2016) The genetics of leaf flecking in maize and its relationship to plant

defense and disease resistance. Plant Physiol 172 (3):1787–1803. https://doi.org/10.1104/pp.15.01870

Olukolu BA, Negeri A, Dhawan R, Venkata BP, Sharma P, Garg A, Gachomo E, Marla S, Chu K, Hasan A, Ji J, Chintamanani S, Green J, Shyu C-R, Wisser R, Holland J, Johal G, Balint-Kurti P (2013) A connected set of genes associated with programmed cell death implicated in controlling the hypersensitive response in maize. Genetics 193(2):609–620. https://doi.org/10.1534/genetics.112.147595

Ordas B, Malvar RA, Santiago R, Sandoya G, Romay MC, Butron A (2009) Mapping of QTL for resistance to the Mediterranean corn borer attack using the intermated B73 × Mo17 (IBM) population of maize. Theor Appl Genet 119(8):1451–1459. https://doi.org/10.1007/s00122-009-1147-6

Pace J, Gardner C, Romay C, Ganapathysubramanian B, Lübberstedt T (2015) Genome-wide association analysis of seedling root development in maize (Zea mays L.). BMC Genomics 16(1):47. https://doi.org/10.1186/s12864-015-1226-9

Pan Q, Li L, Yang X, Tong H, Xu S, Li Z, Li W, Muehlbauer GJ, Li J, Yan J (2016) Genome-wide recombination dynamics are associated with phenotypic variation in maize. New Phytol 210(3):1083–1094. https://doi.org/10.1111/nph.13810

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradbury PJ, Buckler ES (2014) The genetic architecture of maize height. Genetics. https://doi.org/10.1534/genetics.113.159152

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci 108(17):6893–6898. https://doi.org/10.1073/pnas.1010894108

Revilla P, Rodríguez VM, Ordás A, Rincent R, Charcosset A, Giauffret C, Melchinger AE, Schön C-C, Bauer E, Altmann T, Brunel D, Moreno-González J, Campo L, Ouzunova M, Álvarez Á, Ruíz de Galarreta JI, Laborde J, Malvar RA (2016) Association mapping for cold tolerance in two large maize inbred panels. BMC Plant Biol 16(1):127. https://doi.org/10.1186/s12870-016-0816-2

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012a) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217. https://doi.org/10.1038/ng.1033. https://www.nature.com/articles/ng.1033#supplementary-information

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012b) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci 109(23):8872–8877. https://doi.org/10.1073/pnas.1120813109

Rodríguez VM, Butrón A, Malvar RA, Ordás A, Revilla P (2008) Quantitative trait loci for cold tolerance in the maize IBM population. Int J Plant Sci 169(4):551–556. https://doi.org/10.1086/528755

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14. https://doi.org/10.1186/gb-2013-14-6-r55

Samayoa LF, Malvar RA, Olukolu BA, Holland JB, Butrón A (2015) Genome-wide association study reveals a set of genes associated with resistance to the Mediterranean corn borer (Sesamia nonagrioides L.) in a maize diversity panel. BMC Plant Biol 15 (1):35. https://doi.org/10.1186/s12870-014-0403-3

Semagn K, Magorokosho C, Vivek BS, Makumbi D, Beyene Y, Mugo S, Prasanna BM, Warburton ML (2012) Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. BMC Genomics 13(1):113. https://doi.org/10.1186/1471-2164-13-113

Setter TL, Yan J, Warburton M, Ribaut J-M, Xu Y, Sawkins M, Buckler ES, Zhang Z, Gore MA (2011) Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J Exp Bot 62(2):701–716. https://doi.org/10.1093/jxb/erq308

Song Q, Yan L, Quigley C, Jordan BD, Fickus E, Schroeder S, Song B-H, Charles An Y-Q, Hyten D, Nelson R, Rainey K, Beavis WD, Specht J, Diers B, Cregan P (2017) Genetic characterization of the soybean nested association mapping population. Plant Genome 10(2). https://doi.org/10.3835/plantgenome2016.10.0109

Studer AJ, Doebley JF (2011) Do large effect QTL fractionate? A case study at the maize domestication QTL teosinte branched1. Genetics 188(3):673–681. https://doi.org/10.1534/genetics.111.126508

Suwarno WB, Pixley KV, Palacios-Rojas N, Kaeppler SM, Babu R (2015) Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theor Appl Genet 128(5):851–864. https://doi.org/10.1007/s00122-015-2475-3

Thavamanikumar S, Tibbits J, McManus L, Ades P, Stackpole D, Hadjigol S, Vaillancourt R, Zhu P, Bossinger G (2011) Candidate gene-based association mapping of growth and wood quality traits in Eucalyptus globulus Labill. BMC Proc 5(Suppl 7):O15–O15. https://doi.org/10.1186/1753-6561-5-S7-O15

Thirunavukkarasu N, Hossain F, Arora K, Sharma R, Shiriga K, Mittal S, Mohan S, Namratha PM, Dogga S, Rani TS, Katragadda S, Rathore A, Shah T, Mohapatra T, Gupta HS (2014) Functional mechanisms of drought tolerance in subtropical maize (Zea mays L.) identified using genome-wide association mapping. BMC Genomics 15(1):1182. https://doi.org/10.1186/1471-2164-15-1182

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler Iv ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286. https://doi.org/10.1038/90135

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43. https://doi.org/10.1038/ng.746

Van Inghelandt D, Melchinger AE, Martinant J-P, Stich B (2012) Genome-wide association mapping of flowering time and northern corn leaf blight (Setosphaeria turcica) resistance in a vast commercial maize germplasm set. BMC Plant Biol 12(1):56. https://doi.org/10.1186/1471-2229-12-56

Veldboom LR, Lee M, Woodman WL (1994) Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. Theor Appl Genet 88(1):7–16. https://doi.org/10.1007/bf00222387

Warburton ML, Williams WP, Windham GL, Murray SC, Xu W, Hawkins LK, Duran JF (2013) Phenotypic and genetic characterization of a maize association mapping panel developed for the identification of new sources of resistance to Aspergillus flavus and aflatoxin accumulation. Crop Sci 53:2374–2383. https://doi.org/10.2135/cropsci2012.10.0616

Warburton ML, Womack ED, Tang JD, Thrash A, Smith JS, Xu W, Murray SC, Williams WP (2018) Genome-wide association and metabolic pathway analysis of corn earworm resistance in maize. Plant Genome 11. https://doi.org/10.3835/plantgenome2017.08.0069

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, Yan J (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat Commun 5:3438. https://doi.org/10.1038/ncomms4438. https://www.nature.com/articles/ncomms4438#supplementary-information

Wu X, Li Y, Shi Y, Song Y, Zhang D, Li C, Buckler ES, Li Y, Zhang Z, Wang T (2016) Joint-linkage mapping and GWAS reveal extensive genetic loci that regulate male inflorescence size in maize. Plant Biotechnol J 14(7):1551–1562. https://doi.org/10.1111/pbi.12519

Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017) Genome-wide association studies in maize: praise and stargaze. Mol Plant 10(3):359–374. https://doi.org/10.1016/j.molp.2016.12.008

Xiao Y, Tong H, Yang X, Xu S, Pan Q, Qiao F, Raihan MS, Luo Y, Liu H, Zhang X, Yang N, Wang X, Deng M, Jin M, Zhao L, Luo X, Zhou Y, Li X, Liu J, Zhan W, Liu N, Wang H, Chen G, Cai Y, Xu G, Wang W, Zheng D, Yan J (2016) Genome-wide dissection of the maize ear genetic architecture using multiple populations. New Phytol 210(3):1095–1106. https://doi.org/10.1111/nph.13814

Xue S, Bradbury PJ, Casstevens T, Holland JB (2016) Genetic architecture of domestication-related traits in maize. Genetics 204(1):99–113. https://doi.org/10.1534/genetics.116.191106

Xue Y, Warburton ML, Sawkins M, Zhang X, Setter T, Xu Y, Grudloyma P, Gethi J, Ribaut J-M, Li W, Zhang X, Zheng Y, Yan J (2013) Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. Theor Appl Genet 126(10):2587–2596. https://doi.org/10.1007/s00122-013-2158-x

Yan J, Kandianis CB, Harjes CE, Bai L, Kim E-H, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Fernandez MGS, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, DellaPenna D, Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at Zea mays crtRB1 increases [beta]-carotene in maize grain. Nat Genet 42(4):322–327. http://www.nature.com/ng/journal/v42/n4/suppinfo/ng.551_S1.html

Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J (2014) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLOS Genet 10(9):e1004573. https://doi.org/10.1371/journal.pgen.1004573

Yang X, Gao S, Xu S, Zhang Z, Prasanna BM, Li L, Li J, Yan J (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Mol Breed 28(4):511–526. https://doi.org/10.1007/s11032-010-9500-7

Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y, Li J (2010) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet 121(3):417–431. https://doi.org/10.1007/s00122-010-1320-y

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178(1):539–551. https://doi.org/10.1534/genetics.107.074245

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang N, Gibon Y, Gur A, Chen C, Lepak N, Höhne M, Zhang Z, Kroon D, Tschoep H, Stitt M, Buckler ES (2010a) Fine quantitative trait loci mapping of carbon and nitrogen metabolism enzyme activities and seedling biomass in the intermated maize IBM mapping population. Plant Physiol. https://doi.org/10.1104/pp.110.165787

Zhang N, Gur A, Gibon Y, Sulpice R, Flint-Garcia S, McMullen MD, Stitt M, Buckler ES (2010b) Genetic analysis of central carbon metabolism unveils an amino acid substitution that alters maize NAD-dependent isocitrate dehydrogenase activity. PLoS ONE 5(4):e9991. https://doi.org/10.1371/journal.pone.0009991

Zhang X, Warburton ML, Setter T, Liu H, Xue Y, Yang N, Yan J, Xiao Y (2016) Genome-wide association studies of drought-related metabolic changes in maize using an enlarged SNP panel. Theor

Appl Genet 129(8):1449–1463. https://doi.org/10.1007/s00122-016-2716-0

Zhang Z-S, Hu M-C, Zhang J, Liu D-J, Zheng J, Zhang K, Wang W, Wan Q (2009) Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (Gossypium hirsutum L.). Mol Breed 24(1):49–61. https://doi.org/10.1007/s11032-009-9271-1

Zila CT, Samayoa LF, Santiago R, Butrón A, Holland JB (2013) A Genome-Wide Association Study Reveals Genes Associated with Fusarium Ear Rot Resistance in a Maize Core Diversity Panel. G3 Genes Genomes Genet 3(11):2095–2104. https://doi.org/10.1534/g3.113.007328

Part IV

Genomics of Agronomically Valuable Traits

Genomics of Insect Resistance

11

A. Butron, L. F. Samayoa, R. Santiago, B. Ordás and R. A. Malvar

Abstract

The estimate of worldwide annual yield loss in maize due to pests ranges from 7 to 20%. Insects are among the most important pests of maize at all stages of development, from germination to grain filling, and can even be the main spoilers of grains in storage facilities. Each particular insect species shows preference for attacking a specific maize tissue, although other tissues are also susceptible to attack. Insect control has been achieved by diverse approaches, such as the use of insecticides, the modification of cultural practices, the use of biological methods involving parasitoids and sex pheromone-based mate-finding disruption, and the use of resistant cultivars based on monogenic or polygenic resistance. This chapter addresses the current knowledge about genomic regions and genes responsible for maize resistance to insect attack. The information on quantitative trait loci (QTL) and genes involved in resistance has been divided into six sections: (1) soil insects, (2) leaf feeders, (3) stem borers, (4) ear borers, (5) granary pests, and (6) relationships between maize genomics for resistance to insect and agronomical performance.

A. Butron (✉) · L. F. Samayoa · B. Ordás · R. A. Malvar
Misión Biológica de Galicia (CSIC), Apdo. 28, 36080 Pontevedra, Spain
e-mail: abutron@mbg.cesga.es

Present Address:
L. F. Samayoa
Department of Crop Science, North Carolina State University, Raleigh, NC 27695, USA

R. Santiago
Facultad de Biología, Dpto. Biología Vegetal y Ciencias del Suelo, Universidad de Vigo, Campus As Lagoas Marcosende, 36310 Vigo, Spain

R. Santiago
Agrobiología Ambiental, Calidad de Suelos y Plantas (UVIGO), Unidad Asociada a la Misión Biológica de Galicia (CSIC), Vigo, Spain

The estimate of worldwide annual yield loss in maize due to pests ranges from 7 to 20% (Santiago et al. 2016a). Insects are among the most important pests of maize at all stages of development from germination to grain filling and can even be the main spoilers of grains in storage facilities. Each particular insect species shows preference for attacking a specific maize tissue, although other tissues are also susceptible to attack. Root and seedling are preferentially attacked by soil insects (corn rootworms, wireworms, etc.), leaves by sucking (leafhoppers, thrips and aphids) and chewing defoliators (insect specialists and early generations of stem borers), stalks by stem corn borers and ears and kernels by earworms, stem borers, and granary pests (Table 11.1).

Insect control has been achieved by diverse approaches, such as the use of insecticides, the modification of cultural practices, the use of

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_11

Table 11.1 Common name, scientific name, and distribution of most important insect pests of maize

Type	Common name	Scientific name	Distribution
Soil insect	Western corn rootworm (WCR)	Diabrotica virgifera virgifera	In America: from Canada to Central America In Europe: from France to Southern Russia
Soil insect	Northern corn rootworm (NCR)	*Diabrotica barberi*	Northern and Central USA
Soil insect	Mexican corn rootworm (MCR)	*Diabrotica virgifera zeae*	Southern USA, Mexico, Central America
Soil insect	Cucurbit beetle (CB)	*Diabrotica speciosa*	South America
Leaf feeder, ear borer	Fall armyworm (FAW)	*Spodoptera frugiperda*	America and Central and Southern Africa
Leaf feeder	Corn leaf Aphid (CLA)	*Rhopalosiphum maidis*	World wide
Stem borer, leaf feeder	Sugarcane corn borer (SCB)	*Diatraea saccharalis*	Warmer areas of USA, Caribbean, Central America, and warmer areas of South America
Stem borer, leaf feeder	Southwestern corn borer (SWCB)	*Diatraea grandiosella*	USA and Mexico
Stem borer, leaf feeder, ear borer	Asian corn borer (ACB)	*Ostrinia furnacalis*	Asia and Oceania
Stem borer, leaf feeder, ear borer	Spotted stem borer (SSB)	*Chilo partellus*	Southern Asia and Eastern Africa
Stem borer, leaf feeder, ear borer	European corn borer (ECB)	*Ostrinia nubilalis*	Central and Eastern Canada and USA, and Europe
Stem borer, leaf feeder, ear borer	Asian corn borer (ACB)	*Ostrinia furnacalis*	Asia and Oceania
Stem borer, leaf feeder, ear borer	Mediterranean corn borer (MCB)	*Sesamia nonagrioides*	Mediterranean area
Stem borer, leaf feeder, ear borer	African pink stem borer (APSB)	*Sesamia calamistis*	Central and Southern Africa
Ear borer	Corn earworm (CEW)	*Helicoverpa zea*	Temperate and tropical area
Granary pest	Maize weevil (MW)	*Sitophilus zeamais*	Tropical environments throughout the world, less common in temperate zones
Granary pest	Angoumois grain moth (AGM)	*Sitotroga cerealella*	Worldwide
Granary pest	Larger grain borer (LGB)	*Prostephanus truncatus*	Throughout tropics

biological methods involving parasitoids and sex pheromone-based mate-finding disruption, and the use of resistant cultivars based on monogenic or polygenic resistance. In the recent past, the most outstanding contribution of maize breeding for resistance to insects has been the release of "*Bt*" varieties. Transgenic "*Bt* maize" has proved its effectiveness against many insect species, but provokes social rejection in many European countries, and is prohibited in certain agriculture systems like organic agriculture. Besides, recent studies have reported a reduction of efficacy of *Bt* transgenes caused by evolved resistance of some important pests beginning in the late 1990s (Tabashnik et al. 2000), particularly to the earliest commercialized *Bt* crops that produced only

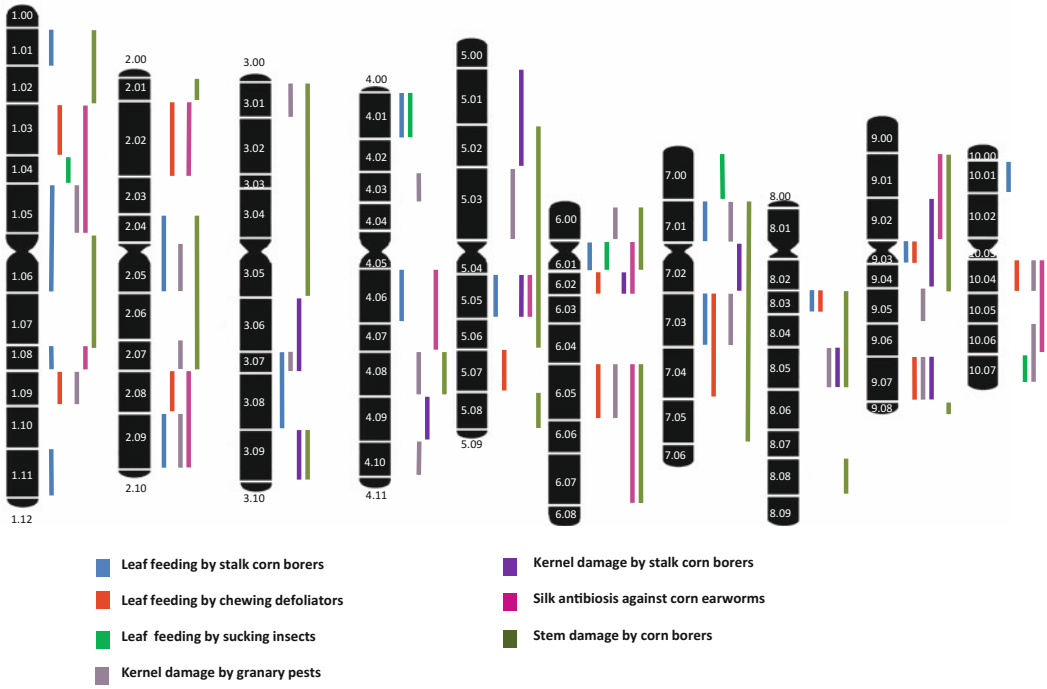

Fig. 11.1 Compilation of maize bins where quantitative trait loci (QTL) for resistance to insects that attack aerial tissues have been located to date. The figure summarizes biparental QTL and genome-wide association SNPs from the following studies: Bohn et al. (1996, 1997), Byrne et al. (1997, 1998), Groh et al. (1998a, b), Khairallah et al. (1998), Lee et al. (1998), Bohn et al. (2000, 2001), Cardinal et al. (2001), Papst et al. (2001), Jampatong et al. (2002), Krakowsky et al. (2004), Papst et al. (2004), Brooks et al. (2005), Cardinal et al. (2005, 2006), Brooks et al. (2007), Krakowsky et al. (2007), Garcia-Lara et al. (2009), Ordas et al. (2009), Garcia-Lara et al. (2010), Li et al. (2010), Ordas et al. (2010), Orsini et al. (2012), Meihls et al. (2013), Samayoa et al. (2014), Betsiashvili et al. (2015), Castro-Alvarez et al. (2015), Foiada et al. (2015), Samayoa et al. (2015a, b), Tzin et al. (2015), Santiago et al. (2016b), Jímenez-Galindo et al. (2017)

one *Bt* toxin where most *Bt* crops today are pyramided with multiple *Bt* toxins (Carriere et al. 2016). The advanced knowledge of genomics of maize natural resistance to pests could solve both problems because research in this discipline will render additional genes to stack in *Bt* varieties in order to slow down the appearance of insect resistance to *Bt* hybrids and will improve varieties for organic and low-input farming (Mohan et al. 2008).

In this sense, this chapter addresses the current knowledge about genomic regions and genes responsible for maize resistance to insect attack. In order to help readers, the information on quantitative trait loci (QTL) and genes involved in resistance has been divided into six sections: (1) soil insects, (2) leaf feeders, (3) stem borers, (4) ear borers, (5) granary pests, and (6) relationships between maize genomics for resistance to insect and agronomical performance. Figure 11.1 summarizes QTL information included in this chapter.

11.1 Soil Insects

There are many species of insects that inhabit the soil and can damage the roots and other subterranean parts of the maize plant. They can be classified, attending to the type of damage on the plant, into: rootworms, cutworms, wireworms, billbugs, webworms, white grubs, corn root aphids, seed-corn maggot, and others (Dicke

1977). No studies on maize genetic resistance to soil insect pests have been reported, except for the rootworms. Ten species or subspecies of the *Diabrotica* (Coleoptera: *Chrysomelidae*) genus are considered corn rootworm pests, with *Diabrotica virgifera virgifera* LeConte (western corn rootworm) being the most economically important species (Hessel 2014) (Table 11.1). Damage to maize plants is mainly done by larvae; newly hatched larvae feed primarily on root hairs and outer root tissue, whereas later larvae instars burrow into the roots to feed. The reduced root system caused by progressive feeding on the roots limits plant ability to take up moisture and nutrients, favors plant lodging, and dramatically reduces yield. Injured roots are also easy entry points for fungi and bacteria that may increase severity of root and stalk rots and premature death.

Corn rootworm species have been reasonably well controlled using a combination of insecticides, transgenic *Bt* crops and agronomical practices, but in recent years these strategies have shown some vulnerabilities, and host-plant resistance is again viewed as an additional tool for controlling maize damage by corn rootworm. The genetic architecture of tolerance (root size and regrowth) and resistance-related (node injury) traits is now being explored, but little can be said about the genomic regions involved in these traits because QTL mapping studies have been performed in populations derived from few crosses. Hessel (2014) reported that chromosome 3 contains genes associated with node-injury resistance; and chromosomes 2, 3, 5, and 7 genes associated with tolerance mechanisms (root size and regrowth size), and suggest that genetic variability for tolerance-related traits is higher than genetic variability for resistance. However, major genes in unexplored genetic regions and/or minor genes were undetected in that study because only a low percentage of genetic variation was explained by the QTL detected. In addition, this author suggests that the genomics of "overcompensation" phenomena (plants performing better under herbivory attack) should be studied because it represents an exceptional target for selection. As carbon reallocation has been proposed as a plant tolerance strategy against root damage by corn rootworms, genomics of carbon reallocation after corn rootworm attack could be the focus of future research (Robert et al. 2014). A more recent study by Bohn et al. (Bohn et al. 2018) reported QTL for root damage ratings in bins 1.07, 2.02, 3.05, 5.03, 6.01, 7.02, 7.06, 8.06, 9.04, and 10.03 using two populations of double haploid lines derived from different resistant and susceptible sources. In this study, the QTL explained between 30 and 50% of the phenotypic variability. The QTL in bin 7.02 was identified in both populations, suggesting that the same gene region may be contributing resistance from independent sources.

11.2 Leaf Feeder Insects

Insects feeding on the leaves of maize remove a portion of the leaves, lacerate the leaves, or remove sap with sucking mouthparts. First-generation stalk borers, armyworms, and beetles remove portions of the leaves by chewing; meanwhile, thrips, aphids, and leafhoppers remove plant juices by sucking. Stem borers, armyworms, and grasshoppers (*Melanoplus spp*.) feed on the young leaves and whorls of maize plants, producing small holes and notchings on the leaf margins. Heavy infestations of these pests can consume the whole leaf except the tougher part, the leaf midrib; photosynthesis is reduced, and, consequently, growth of the maize plant is inhibited and yield is reduced. Thrips (*Frankliniella spp.*) are commonly found in leaf whorls, ears, and on the underside of developed leaves; when they feed on young seedlings, plants get stunted. A common sign of heavy infestation by thrips is distorted leaves that turn brownish around the edges and cup upward, but the actual thrips injury does little damage and usually the plants will recover. Aphids (*Rhopalosiphum maidis* or corn leaf aphid (CLA), *Rhopalosiphum padi*, and *Schizaphis graminum*) insert their stylets into phloem sieve elements causing mottling and discoloration of the leaves and, when colonies are large, abnormal synthesis of anthocyanin resulting in the

reddening of the leaves. Infested plants become covered with sweet, sticky honeydew secretions, and mold fungi grow causing interferences with photosynthesis and kernel development. Finally, leafhoppers (*Dalbulus maidis*) suck out juices and transmit the corn stunt disease (caused by the bacteria *Spiroplasma kunkelii*) that is much more debilitating to the plants than the direct feeding damage caused by the insect. Most of the studies screening for genomic regions involved in maize resistance to leaf feeding insects are focused on chewing insects, with fewer results on sucking insects and no studies on thrips and leafhoppers.

Stem borers feed on leaf tissues when infesting maize plants at the vegetative stage or during the initial instar stages of larval development on plants at the reproductive stage. Several studies have tried to locate genomic regions involved in resistance to leaf feeding by stem borers such as *Ostrinia nubilalis* (European corn borer, ECB), *Ostrinia furnacalis* (Asian corn borer, ACB), *Diatraea grandiosella* (southwestern corn borer, SWCB) and *Diatraea saccharalis* (sugarcane corn borer, SCB), but no information is available about other important stem borers like *Sesamia nonagrioides*, *Sesamia calamistis,* or *Chilo partellus*.

11.2.1 European and Asian Corn Borers

Maize leaf feeding damage by *Ostrinia* species has been assessed using a subjective 9-point visual scale, and studies in diverse mapping populations have detected genomic regions containing QTL with significant additive effects on resistance to leaf feeding damage on all chromosomes, in bins 1.01, 1.05–1.06, 1.08, 2.04, 2.05, 1.11, 2.09, 3.08, 4.01, 4.06, 6.01, 5.04–5.05, 7.01, 7.03, 8.03, 9.03 10.01 (Jampatong et al. 2002; Cardinal et al. 2006; Li et al. 2010; Orsini et al. 2012). In general, these QTL had small effects and explained low percentages of phenotypic variability, except those located in bin 4.01 which, in some genetic backgrounds, explained more than 40% of phenotypic variability. As the region 4.01 includes several *Bx* loci of the benzoxazinoid pathway and benzoxazinoid compounds have proved an antibiotic effect against ECB larvae, polymorphisms at structural or regulator genes of this pathway can have an important impact on resistance to leaf feeding (Wouters et al. 2016). QTL with significant dominance effects on resistance to leaf feeding by ECB have been found, as well as significant epistatic interactions between QTL (Jampatong et al. 2002; Cardinal et al. 2006). The genomics of resistance to leaf feeding by *Ostrinia* species is far from being elucidated because a small proportion of maize genetic diversity (just four biparental populations) has been explored; QTL detected in some studies explained a low proportion of genetic variation, and we lack appropriate tools to uncover epistatic effects.

11.2.2 Sugarcane and Southwestern Corn Borers

Diverse maize germplasm has been developed through selection for reduced leaf feeding damage by the first generations of SCB and SWCB. Resistance to SWCB is polygenic and involves primary additive genetic variation (Scott and Davis 1978; Williams et al. 1989; Thome et al. 1992). QTL studies confirmed these results (Bohn et al. 1996, 1997; Groh et al. 1998a; Khairallah et al. 1998). Moreover, some of resistance factors to SWCB appear to confer resistance to other insect species, including sugarcane borer SCB and other lepidopteran species such as ECB or *Helicoverpa zea* (Thome et al. 1992; Bohn et al. 2001; Cardinal et al. 2001; Brooks et al. 2007).

Bohn et al. (1997) identified 11 QTL (bins 1.06, 1.07, 1.11, 3.05, 5.06, 5.07, 7.02, 7.04, 9.05, 10.03–10.04) affecting resistance to leaf feeding by SCB and SWCB in 171 F_3 families from the cross CML131 (susceptible) × CML67 (resistant). While the majority of the QTL for resistance to both insect species co-localized in this population, only two or three were detected in the same positions in a second population derived from the cross between Ki3 (susceptible)

and CML139 (resistant) (Bohn et al. 1997). Khairallah et al. (1998) and Groh et al. (1998a, b) further extended these mapping QTL studies for resistance to leaf damage by borers to recombinant inbred lines (RILs) derived from the same crosses, CML131 × CML67 and Ki3 × CML139. The QTL were more consistent across populations (RILs vs. $F_{2:3}$) derived from CML131 × CML67 (QTL for RILs detected in bins 1.01–1.02, 1.03, 1.06, 1.10, 5.05–5.06, 7.03, 8.05–8.06, 8.06–8.07, 9.02–9.04) than across those derived from Ki3 × CML139 (QTL for RILs detected in bins 3.07–3.09, 5.05–5.07 (2 QTL), 6.04–6.05, 6.06, 8.02, 9.05), but this could be a consequence of sampling effects that lead to detection of different sets of QTL rather than to the relevance of the dominance gene action. These results suggested that although a shared genetic basis of resistance against leaf feeding by both borers can exist, additional independent mechanisms of defense against both species should be also considered. One QTL on chromosome 9 (bin 9.05) was detected for resistance to both borers using $F_{2:3}$ mapping populations developed from crosses CML131 × CML67 and Ki3 × CML139 (Bohn et al. 1997). In this region of chromosome 9, the gene *brittle stalk 2* (*bk2*) causes susceptibility to easy leaf breakage by affecting the stiffness and toughness of leaves. Bergvinson et al. (1995) already found that the consumption rate of ECB larvae was negatively correlated with leaf toughness, and Ching et al. (2006) demonstrated that the mutant *bk2*, exhibits dramatically reduced tissue mechanical strength because the mutation interferes with the deposition of cellulose in the secondary cell walls.

Finally, Willcox et al. (2002) identified three QTL regions (bins 7.04, 9.03, 10.03–10.06) related to SWCB resistance in a mapping population derived from the cross between the resistant inbred CML67 (used by Bohn and Groh above) and the susceptible inbred CML204. A QTL on chromosome 9 in or near bin 9.03 was identified. The *gl15* locus, demonstrated to play a role in vegetative phase change from juvenile to adult phase in maize leaves, was proposed as a candidate gene (Moose and Sisco 1996).

11.2.3 Fall Armyworm

FAW feeding in the whorl often produces a characteristic row of perforations in the leaves, while older larvae cause extensive defoliation, often leaving only the midribs and stalks of corn plants, causing a ragged appearance. In a study estimating combining ability for resistance to FAW and SWCB, Williams et al. (1989) observed a strong correlation of general combining ability (GCA) for leaf feeding, larvae weights, and larvae number between the two pests. The authors suggested that selection for resistance to one insect pest could improve resistance to the other. Further studies suggested that vegetative phase change, which is controlled by *gl15* gene, is a primary mechanism affecting resistance to both first generations of FAW and SWCB (Williams et al. 1998; Williams and Davis 2000).

In later studies combining leaf damage by SWCB and FAW, Brooks et al. (2005) identified 7 QTL for both SWCB (bins 1.04, 1.11, 5.02, 6.02, 7.02–7.03, 9.03, 10.04) and FAW (bins 1.09, 2.08, 6.02, 7.04, 8.03, 9.03, 10.04) resistance in $F_{2:3}$ families from the cross Mo17 (susceptible) × Mp704 (resistant). Estimated genotypic variance explained by each QTL ranged from 3 to 23% for resistance to SWCB and from 5 to 36% for resistance to FAW. Loci on chromosomes 6, 9, and 10 appeared to affect leaf feeding damage ratings under infestation with both SWCB and FAW, exhibited similar effects, and had similar gene action on both traits. In a subsequent study, Brooks et al. (2007) identified 4 and 7 QTL for resistance to SWCB (bins 3.08, 5.04, 6.05, 9.05) and FAW (bins 1.03, 2.02, 5.07, 6.05, 7.03, 9.03, 9.07), respectively, in $F_{2:3}$ families derived from the cross A619 (susceptible) × Mp704 (resistant). The genetic model best fitting the data for SWCB and FAW explained just 39% and 48% of the phenotypic variance, respectively. When a multiple trait analysis was performed using both mapping populations, A619 × Mp708 and Mo17 × Mp704, they identified 15 QTL related to leaf damages by both pests (bins 1.05, 1.02, 1.11, 2.02, 5.02, 5.04, 5.07, 6.05, 7.02, 7.03, 7.04, 8.03, 9.03, 9.05,

10.04) showing that many genomic regions contain genes conferring resistance to both insects. Loci detected tended to be simply inherited in an additive or dominant manner, although significant epistatic interactions were also found. In relation to candidate genes within the QTL intervals, they proposed once again *gl15* on chromosome 9, involved in the transition of vegetative phases; *p1* on chromosome 1, involved in the maysin pathway, as maysin is a flavone glycoside contained in the silks of maize varieties resistant to FAW (Waiss et al. 1979; Wiseman and Widstrom 1992); or *mir* (maize insect resistant) genes that encode cysteine proteinases involved in the ability of callus from Mp704 and Mp708 to retard FAW larval growth (Pechan et al. 2000). In addition, microarray and proteomic analyses of insect-infested plants allowed the identification of proteins with probable anti-herbivore activity. Fescemyer et al. (2013) identified 1-cysteine protease (*Mir1-CP*, in bin 6.02) affecting the peritrophic matrix of FAW, and Chuang et al. (2014) a ribosome-inactivating protein 2 (*RIP2*, in bin 7.04) with a not yet determined activity in the midgut of FAW, but within intervals of major QTL for caterpillar resistance (Brooks et al. 2005, 2007).

11.2.4 Corn Leaf Aphid

The main damage of CLA is on leaves, although all aboveground parts of maize plant are susceptible. Infestation at the seedling stage slows development and reduces plant height; tassel infestations can prevent pollen shed due to the accumulation of sticky honeydew (Carena and Glogoza 2004). Additionally, CLA can transmit plant viruses, including *Maize dwarf mosaic virus*. Resistance to CLA in maize was first reported by Gernert (Gernert 1917) in F_1 hybrids of *Zea mays* ssp. *mexicana* (synonymous to *Euchlaena mexicana*) × yellow dent maize. Since that time, some studies indicated that resistance was monogenic (Chang and Brewbaker 1976; Lu and Brewbaker 1999), whereas others showed that multiple genes with additive effects contributed to resistance (Long et al. 1977; Bing and Guthrie 1991; Bing et al. 1992). In one series of experiments, aphid resistance was recessive, and further analysis identified two resistance loci, *aph1* on chromosome 10 (bin10.07) and *aph2* on the short arm of chromosome 2 (bin 2.02), (Chang and Brewbaker 1976; Lu 1999; Lu and Brewbaker 1999; So 2003; So et al. 2010).

More recently, large variation for resistance to CLA was found among parental lines of the maize nested association mapping (NAM) population (Meihls et al. 2013) and genetic mapping for CLA resistance was conducted using RILs derived from crosses between B73 and the most aphid-sensitive inbred lines of the NAM (CML 52, CML69, CML277, and CML322). A natural transposon insertion in *Bx10c* (GRMZM2G023325 on chromosome 1) was associated with increased resistance. This gene encodes a methyltransferase that converts DIMBOA-Glc (2-(2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one)-beta-D-glucopyranose) to HDMBOA-Glc (2-(2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one)-beta-D-glucopyranose). In vitro, HDMBOA-Glc was more toxic to CLA than DIMBOA-Glc, but the decline of DIMBOA-Glc upon methylation to HDMBOA-Glc was associated with reduced callose deposition that is an aphid defense response in vivo (Meihls et al. 2013; Ahmad et al. 2011). Additional genes involved in resistance to CLA have been looked for in the IBM (B73 × Mo17) population because both parental inbreds carry the transposon insertion in the *Bx10c* gene, although B73 is sensitive and Mo17 is resistant (Betsiashvili et al. 2015). These authors identified QTL on chromosomes 4 and 6 that accounted for 15% and 27% of the total genetic variance for resistance to aphids, respectively. A regulator of *Bx1* expression could probably be behind the QTL on chromosome 4 because

DIMBOA is required for callose induction in maize (Ahmad et al. 2011), a DIMBOA QTL in the maize IBM population was fine-mapped to a specific *cis*-regulatory region more than 100 kbp upstream the *Bx1* gene, and expression of *Bx1* gene in developing Mo17 seedlings showed longer persistence compared to expression in B73 (Zheng et al. 2015). However, the QTL on chromosome 6 could be related to a resistance mechanism other than DIMBOA accumulation because the QTL interval for aphid resistance contains neither known benzoxazinoid biosynthesis genes nor genetic evidence for a DIMBOA-related QTL.

QTL for CLA progeny production were also studied in a RIL population derived from B73 × Ky21 which was pre-infested with beet armyworm (BA; *Spodoptera exigua*) larvae (Tzin et al. 2015). Three QTL for progeny production were found on chromosomes 1, 7, and 10. The genomic region on chromosome 1 that affects caterpillar-induced aphid resistance (resistance to aphids in plants previously challenged with BA larvae) contains a cluster of three benzoxazinoid O-methyltransferase genes that convert DIMBOA-Glc to HDMBOA-Glc, whereas QTL on chromosomes 7 and 10 could be related to a defense mechanism different from DIMBOA-Glc mediated resistance.

Recently, another metabolite likely implicated in resistance to CLA has been discovered: the non-protein amino acid 5-Hydroxynorvaline (Yan et al. 2015). This compound, when added to aphid artificial diet at concentrations similar to those found in maize leaves and stems, reduces CLA reproduction and accumulation of this compound increases after herbivory, as well as in response to treatments with the plant signaling molecules methyl jasmonate, salicylic acid, and abscisic acid (Yan et al. 2015). The joint linkage analysis of RILs derived from B73 crossed to CML103, CML228, CML277, and Ky21 identified significant QTL for methyl jasmonate-induced 5-hydroxynorvaline content on chromosomes 5 and 7, which explained 16 and 20% of the phenotypic variance, respectively.

11.2.5 Benzoxazinoids Effects on Leaf Feeder Insects

Benzoxazinoids (BXs) are well-known maize defensive metabolites effective against a large number of herbivores and pathogens (Niemeyer 2009) and have been extensively studied. However, DIMBOA concentrations decrease as the plant develops, so this mechanism fails to protect plants at tasseling and later stages of development (Mihm 1985). BXs are a class of indole-derived plant chemicals comprising compounds with a 2-hydroxy-*2H*-1,4-benzoxazin-3(*4H*)-one skeleton and their derivatives. BXs are commonly stored as glucosides (-Glc) in vacuoles of undamaged plant cells, and hydrolysis by β-glucosidases increases their reactivity and toxicity (Frey et al. 2009; Niemeyer 2009). The formation of BXs as well as their genetics and evolution in plants has been comprehensively reviewed (Gierl and Frey 2001; Frey et al. 2009; Makowska et al. 2015). In summary, the core biosynthetic pathway for DIMBOA-Glc in maize is encoded by eight genes (*Bx1–Bx8*) that are tightly linked at the top of maize chromosome 4. Four additional genes, *Bx9* that encodes a glucosyltransferase, and *Bx10, Bx11, and Bx12* which encode methyltransferases involved in conversion of DIMBOA-Glc to the more toxic HDMBOA-Glc (Oikawa et al. 2004; Glauser et al. 2011), are located on chromosome 1 (Gierl and Frey 2001; Ahmad et al. 2011; Meihls et al. 2013). More recently, Handrick et al. (2016) identified *Bx13* and *Bx14*, which participate in the conversion of DIMBOA-Glc and DIM_2-BOA-Glc to TRIMBOA-Glc and HDM_2BOA-Glc, respectively, on chromosome 2. Furthermore, two genes encoding for glucosylglucosidases, *Zmglu1* and *Zmglu2,* are located on chromosome 10. In addition, QTL likely containing regulatory elements of the DIMBOA pathway have been detected in bins 3.08, 6.01, 6.05, 7.02, and 8.06 (Butrón et al. 2010).

As previously mentioned, QTL for resistance to leaf feeders have been located in regions containing genes of the benzoxazinoid pathway and variability for resistance may be related to

natural variation for benzoxazinoid content. Control of benzoxazinoid content has already been proven for the aphid resistance QTL in bin 1.04 that co-localized with a QTL for HDMBOA-Glc and seems to be consequence of natural variation at gene *Bx12* (Jampatong et al. 2002; Krakowsky et al. 2004; Cardinal et al. 2006; Li et al. 2010; Meihls et al. 2013). However, other QTL for benzoxazinoid content did not have any significant effect on resistance to leaf feeders and could be consequence of unknown maize defense responses that are linked in repulsion with genes involved in the benzoxazinoid pathway (Betsiashvili et al. 2015). Therefore, defense mechanisms involved in resistance to leaf feeders did not seem restricted to benzoxazinoid accumulation, in part, due to coevolution of pests and hosts. For example, specialist herbivores, such as FAW, can be insensitive to variation in benzoxazinoid content or have specific DIMBOA detoxification enzymes (Glauser et al. 2011). Thus, this species might even be attracted to plants with higher BXs content. In addition, within the genera *Spodoptera*, FAW as a specialized grass feeder seems to be more resistant to BXs than the generalists *S. littoralis*, *S. eridania,* and *S. exigua* (Wouters et al. 2016).

11.3 Stem Borer Insects

There is a long list of insects that feed on the maize stem and belong to the group of corn stem borers (Table 11.1). Corn stem borers usually spend most of their larval stage within the stem, feeding on the pith, a tissue with a high concentration of nutrients. Depending on environmental conditions and insect species, larvae could move from the stem bottom to the tassel creating long galleries which could drastically affect kernel production and stem integrity.

The length of tunnels made by borers in the stem has been commonly used to estimate the level of resistance/susceptibility of genotypes. Sources of resistance have been identified for different species of corn stem borers (Williams et al. 1983; Barrow 1985, 1987; Hawk 1985; van Rensburg and Malan 1990; Malvar et al. 1993; Sekhon and Kanta 1994; Williams and Davis 1994; Kumar 1994a, b; van Rensburg and van den Berg 1995; Barry et al. 1995; Kumar and Mihm 1996; Schulz et al. 1997; Melchinger et al. 1998; Bohn et al. 2000; Ajala et al. 2001; Butron et al. 1999a). Maize resistance to corn stem borers is polygenic with moderate to high heritability (0.5–0.8) (Schulz et al. 1997; Cardinal et al. 2001; Krakowsky et al. 2004; Ordas et al. 2009; Samayoa et al. 2014, 2015a, b; Foiada et al. 2015), with additive effects being the most important genetic effects for tunnel length (Butrón et al. 1998, 1999b; Cartea et al. 2001). Many QTL experiments for stem tunneling by stem borers have been reported in maize, but only genetic regions involved in resistance to ECB and MCB have been localized.

11.3.1 European Corn Borer and Mediterranean Corn Borer

The location of detected QTL for resistance to stem tunneling by borers is shown in Fig. 11.1. Most mapping populations consisted of biparental populations derived from a cross between a susceptible line and a resistant line. The resistant lines used to map QTL for tunnel length belong to diverse germplasm groups: European dent (D06), corn belt dent (Mo17, A509, De811, B52), European flint (EP39), and tropical (Mo47, CML103). The level of resistance of the resistant lines is also variable, from highly resistant, for example, De811 and Mo47, to moderate, for example, Mo17. The number of QTL detected in each experiment can vary due to differences in: resistance between parental lines, number of families, marker coverage in the mapping population, and precision of phenotyping. However, the number of QTL detected was relatively similar in all ECB experiments: about 9 QTL per experiment. Also, the number of QTL detected was similar in all MCB experiments, but the number of QTL detected was about a third of those detected in ECB experiments. As the damage produced by ECB larvae is lower than

that produced by MCB larvae, genotype differences for resistance would be less conspicuous under infestation with MCB larvae (Sandoya et al. 2008).

The QTL for resistance to stem tunneling are randomly distributed across the genome on all chromosomes. No QTL for resistance to stem tunneling by MCB were found on chromosome 2, although multiple QTL were found for resistance to tunneling by ECB on that chromosome. No more than 2 or 3 QTL co-localized in the same bin across mapping populations in spite of the large number of populations, suggesting that there is not clear evidence of QTL hot spots for resistance to stem borers. The proportion of genotypic variance explained by individual QTL was usually not large, even when parental lines showed obvious contrasting levels of resistance, and in spite of the low number of families evaluated (about 200) which results in overestimation of the QTL effect. The total percentage of genotypic variance for resistance to stem tunneling by ECB explained by QTL was around 40–60% (Bohn et al. 2000; Cardinal et al. 2001; Krakowsky et al. 2004), while lower percentages of genotypic variances for resistance to MCB were explained by the QTL detected (10–30%) (Ordas et al. 2009, 2010; Samayoa et al. 2014, 2015a, b).

Therefore, the main conclusion of the multiple experiments with biparental mapping populations is that resistance to stem tunneling by corn borers is controlled by multiple loci of small effect distributed randomly along the genome and most of the genetic variation was undetected in the individual experiments. Validation of QTL identified in those experiments has to be carried out before using them in marker-assisted selection programs and/or performing fine-mapping studies for cloning the genes underlying the QTL. Our group developed a set of heterogeneous inbred families (Tuinstra et al. 1997) to validate a QTL for resistance to stem tunneling that co-localized with a QTL for yield in a previous study conducted with a biparental population (Jiménez-Galindo 2017). The QTL for yield was successfully confirmed, but not the QTL for resistance to stem tunneling probably due to the fixation of some genomic regions during the development of the heterogeneous inbred families. The limited precision of QTL localization using biparental mapping populations (mainly due to sample size), the small effects of the QTL, and the lack of consistency across different germplasms restrict the use of the QTL in marker-assisted selection. In spite of this, Flint-García et al. (2003) have empirically demonstrated that marker-assisted selection can be effective in selecting for resistance to tunneling by ECB. The efficiency of marker-assisted selection can be considerably increased by the use of connected biparental populations and genome-wide prediction (Foiada et al. 2015).

A recent genome-wide association study for maize resistance against MCB was performed using a maize diversity panel of 282 inbred lines representing a high proportion of maize genetic variability available in the public breeding sector around the world (Liu et al. 2003; Flint-Garcia et al. 2005; Samayoa et al. 2015b). A set of nine genes containing or adjacent to the most significant single nucleotide polymorphism (SNP) markers were proposed as candidates for maize resistance to stem tunneling by MCB. Some candidate genes reported in that study are part of signaling pathways, and other act as regulators of expression under biotic stresses and other encode proteins that can affect the digestive system of the insect. Although these genes should be also validated, this kind of study provides a more accurate picture about the genetic architecture of maize resistance against stem borers because a large percentage of maize genetic variability has been evaluated.

Independent of the mapping population, the precise mapping of QTL related to stem tunneling is limited by difficulties in phenotyping. Artificial infestation with corn borer eggs or larvae is usually carried out to guarantee the same insect pressure on each family of the mapping population. However, other factors

could impede a homogeneous spread of the pest across genotypes: differences for voraciousness and mortality across individuals of the insect population, random natural infestation that is added to artificial infestation, etc. The large experimental errors inherently associated with stem tunneling by corn borers together with its highly quantitative nature hamper progress elucidating the genetic variants responsible for the phenotypic variation. Higher precision in phenotyping with more control of environmental variables will be needed to deepen in the genetic architecture of resistance to stem borers.

Behind almost all defense mechanisms against insect pest are complex biochemical processes (McMullen et al. 2009). Several studies have focused on identification of genetic factors that affect the biochemistry involved in maize defense against insects. As high values of crude fiber had been associated with maize resistance to stem borers (Hedin et al. 1993), Papts et al. (2004) tried to locate QTL for crude fiber and resistance to stem tunneling in the same population, but QTL for both traits did not co-localize. However, evidence of the important role of maize cell wall in maize defense against insects has been shown (Barros-rios et al. 2011) and clustering of QTL for resistance to stem borers and for cell-wall components has been observed in other populations (Cardinal and Lee 2005; Krakowsky et al. 2007; Santiago et al. 2016b).

11.4 Ear Borer Insects

The corn earworm (Table 11.1) can be considered a true ear borer because they prefer feeding on the tips of maize ears over any other maize tissue. But there are other insect pests that can damage the ear even though they prefer other tissues (Dicke 1977; Hardwick 1965). For example, stem corn borers can inflict damage on kernels as well as fall armyworms (Dicke 1977; Overholt et al. 1994; Butron et al. 1998). In general, there are few studies addressing the genomics involved in ear resistance to insect attack and those have focused in just two species, CEW and MCB.

11.4.1 Mediterranean Corn Borer

The MCB is a stem borer, but larvae of this lepidopteran species can secondarily damage maize kernels; the kernel is accessed by larvae through husks or more often through the shank (Velasco et al. 2002). Most studies to uncover the genomic regions involved in resistance to ear attack (based on visual ratings of proportion of kernels damaged) by MCB have been done in biparental populations expressing low genetic variability for this trait. In those mapping studies, few QTL for kernel resistance were discovered and were located in bins 2.05–2.06, 5.00–5.01, 5.04–5.05, 8.05, 9.02–9.04, and 9.07 (Ordas et al. 2009; Santiago et al. 2016b, Samayoa et al. 2014, 2015a; Jímenez-Galindo et al. 2017). However, in a later study, genomic regions controlling resistance of the maize ear to MCB attack were explored in the same panel of 282 inbred lines used for studying genome-wide associations between SNPs and resistance to stem tunneling by MCB (Liu et al. 2003; Flint-Garcia et al. 2005; Samayoa et al. 2015b). The genome-wide association study performed by Samayoa et al. (2015b) revealed that genes involved in kernel resistance to MCB could be also found in nine genomic regions of chromosomes 3, 4, 5, 6, and 7. Each region accounted for 6–8% of the phenotypic variability, and 10 genes located in these regions were proposed as candidate genes; these genes would have relevant roles in reactive oxygen species (ROS) scavenging, disturbance of the digestive process of insects, and plant hormone signaling.

11.4.2 Corn Earworm

CEW larvae firstly feed on the silks, but, when the husks are loose and/or the silk channel is short, they easily access to the tip of the ear and feeding extends on maize kernels (Dicke 1977). No studies on genomic regions involved in kernel resistance to feeding by CEW have been done; research to date was focused on silk antibiosis and, specifically, on genetic variation for known silk antibiotic compounds. Antibiosis

of silks was assessed in bioassays with neonate larvae of CEW fed with meridic diets containing silks or methanol extracts from the silks in which the effects on larval weights were measured (Wiseman et al. 1992). QTL for maize silk antibiosis were found in bins 1.03–1.05, 1.08, 2.02, 2.08–2.09, 4.06–4.07, 5.05, 6.01–6.02, 6.05–6.07, 9.01–9.02, and 10.04–10.06 (Byrne et al. 1997, 1998; Lee et al. 1998). Resistance to ear damage by CEW has been associated with high levels of phenylpropanoid compounds that act as antibiotic substances against CEW larvae (Snook et al. 1997). However, high antibiotic concentration must be accompanied by good husk coverage (both husk length and tightness) of the ear that slows down the movement of larvae through the silk channel and increases the exposure of larvae to the antibiotic substance (Butron et al. 2001). Known silk antibiotic compounds against CEW larvae include the phenylpropanoid chlorogenic acid and several glycosyl flavones derived from the flavonoid branch of the phenylpropanoid pathway, such as maysin, apimaysin, 3'-methoxymaysin, isoorientin, and 4"-hydroxy-maysin (Snook et al. 1997). In populations segregating for antibiosis against CEW and maysin (the predominant glycosyl flavone compound) content, there was not total agreement between QTL for both traits and, more surprisingly, regions with high impact on maysin did not affect antibiosis (McMullen et al. 1998).

Therefore, although many studies have been addressed the study of genomics involved in the synthesis of maysin and other antibiotic compounds, the complexity of maize genetic resistance to CEW damage is far from being totally elucidated. Diverse defense mechanisms are likely involved and may not be completely independent; for example, linkage repulsion between favorable alleles for ear tightness and maysin content was observed as well as between favorable alleles for two different antibiotic compounds, chlorogenic acid and maysin (Byrne et al. 1996, 1998; McMullen et al. 2004; Lee et al. 1998; McMullen et al. 1998, 2001, Guo et al. 1999, 2001; Butron et al. 2001; Bushman et al. 2002; Szalma et al. 2002, 2005; Zhang et al. 2003; Cortes-Cruz et al. 2003; Meyer et al. 2007; Sharma et al. 2012). However, knowledge of the genomic control of maysin accumulation in silks sheds light on: (1) the relative importance of genetic variation at regulatory and structural genes in generating variability for specific metabolites involved in resistance; (2) how genes involved in other branches of the main biosynthesis pathway could also play relevant roles in metabolite accumulation because both branches of the pathway share intermediates; (3) the existence of independently regulated complexes that allow for independent synthesis of similar compounds; (4) how more than one gene could be underlying a QTL; (5) the importance of epistatic interactions between major and other genes; and (6) how conditioning for epistatic factors greatly increases the power to detect secondary QTL. Most studies addressing maize genetic variability for maysin content in silks were performed in biparental populations. However, genetic variability in four candidate genes (*p*, *pericarp color*, a myb transcription factor homolog*; chalcone synthase* genes, *c2* and *whp1*; and *NADPH dehydroflavonol reductase*, gene *a1*) among 86 inbred lines was explored for their influence on phenotypic variability for maysin content (Szalma et al. 2005). These authors found that a model including main and epistatic (locus × locus and locus × background) effects for genetic variability for three genes (*p* at 1.03, *a1* at 3.09, and *whp1* at 2.09) only accounted for 36% of phenotypic variability for maysin.

Fewer studies have been devoted to identifying genomic regions involved in husk morphology traits related to CEW kernel damage (Butron et al. 2001; Cui et al. 2016). However, genetic variability for husk morphology has been studied in an array of 508 inbred lines and some general considerations about the genomics of resistance-related husk characteristics can be outlined (Cui et al. 2016). Although heritability estimates for husk leaf number, length, width, and thickness ranged from 0.49 to 0.75, few significant associations between SNP polymorphisms and phenotypic variation were found and collectively only explained 11.2, 0.0, 4.9, and 21.4 of phenotypic variance, respectively. Therefore, many

genes of minor effect appear to contribute to genetic variation for husk characteristics that could affect maize resistance to corn earworm.

11.5 Granary Pests

Maize problems do not end when the grain is harvested; insects attack the grain during storage, causing worldwide losses of around the 20% or more of stored grain, especially in developing countries (Bergvinson and Garcia-Lara 2004; Midega et al. 2016). Insects considered as primary granary pests cause damage to whole grains by directly feeding on them, while secondary pests attack only broken grain, dust, and milled products. There are several insect species that attack stored grain (Table 11.1), but we will focus on three: maize weevil, angoumois grain moth, and larger grain borer. Their behavior, distribution, and control methods are known, but few studies about genetic resistance to these pests have been carried out. Consequently, few breeding programs have focused on improving resistance to grain storage pests and the search for associations between regions of maize genome and resistance has been restricted to resistance to MW.

Few resistance sources to LGB were identified among Caribbean germplasm (Kumar 2002), but considerable variation was found among maize cultivars and populations for resistance to this pest (Arnason et al. 1992; Mwololo et al. 2012). In addition, resistance to this pest has been successfully improved through breeding (Tefera et al. 2016). So, the genetic component of phenotypic variability for resistance to LGB is relevant. However, no research has been carried out to locate QTL associated with resistance to LGB, although the inheritance of this resistance seems to be quantitative.

Despite the importance of the AGM attack, few specific studies have been focused on resistance to AGM. Ahmed et al. (2013) evaluated the resistance of maize varieties to AGM in choice and no choice bioassays and found varieties with partial resistance to AGM. Genetic differences among maize landraces for kernel infestation rate by AGM in the field were also tested because AGM larvae, unlike other granary pests, can infest maize kernels in the field before harvest (Butron et al. 2008). Nevertheless, nothing is known about genetics or genome regions associated with AGM resistance.

11.5.1 Maize Weevil

Sources of resistance to MW have been identified, and possible mechanisms involved in resistance have been proposed: biophysical grain characteristics such as kernel hardness, vitreosity and density or pericarp proportion in the kernel; crude fiber, trypsin inhibitor, and phenolic acid contents; and antioxidant capacity (Dobie 1974; Arnason et al. 1994; Garcia-Lara and Bergvinson 2014; Nwosu 2016). Inheritance of maize resistance to MW seems to be complex but with an important additive component, and regions of genome related to resistance to WM have been identified (Dhliwayo and Pixley 2003; Dhliwayo et al. 2005; Garcia-Lara et al. 2009; Dari et al. 2010; Castro-Alvarez et al. 2015).

QTL involved in MW resistance were investigated using two biparental mapping populations. The first was a $F_{2:3}$ population derived from a tropical yellow maize population; meanwhile, the second was a RIL population developed from the cross between inbreds derived from Tanzanian and Caribbean germplasm. Grain damage, grain weight losses, Dobie index [calculated from the formula (logP × 100)/MDT, where P is the adult progeny and MDT is the median development time], and adult progeny of MW were used to estimate resistance to MW attack in the first study, while grain weight losses, flour production by MW, and adult progeny were used in the second study. A total of 21 QTL, distributed on all chromosomes (bins 1.05, 1.09, 2.05, 2.07, 2.09, 3.01, 3.07, 4.03, 4.08, 4.10, 5.03, 6.00, 6.05, 7.01, 7.03, 8.05, 9.05, 9.07, 10.04, 10.06, and 10.07), were found for resistance traits. Regions 4.08, 7.01, and 10.04 were interesting because QTL for three resistance traits overlapped, but especially the QTL in bin 10.07 in which QTL for four resistance traits

were found. Only one QTL (bin 10.04) was found in both mapping populations. This indicates the need to extend the study to new populations to delve into the genome regions responsible for MW resistance. Phenotypic variance explained by all putative QTL affecting the trait ranged from 11 to 51% (Garcia-Lara et al. 2009; Castro-Alvarez et al. 2015).

Three QTL for grain hardness localized to regions or nearby regions containing QTL for MW resistance (bin 7.02, 7.04, and 10.06), while one QTL (bin 3.07) for pericarp/grain ratio co-localized with a QTL for Dobie index (Garcia-Lara et al. 2009). QTL for some of the principal phytochemicals involved in mechanical fortification of pericarp cell wall co-localized with QTL for resistance: QTL for ferulic acid were found in bin 2.09, 301, 3.07, 5.03, 6.05, and 10.06; for *p*-coumaric acid in bin 1.09 and 9.07; for diferulic acids in bin 4.08 and 6.05; for total phenolic acids in bin 3.01 and 10.06; for structural proteins like a hydroxyproline-rich glycoprotein in bin 8.05. Likewise, QTL for biochemical compounds implicated in antibiosis, such as phenolic acid amides, were located in regions containing QTL for maize resistance to MW: QTL for *p*-coumaroyl-feruloyl putrescine were found in bin 5.03, 3.01, 6.05, and 10.06 and for diferuloyl putrescine in bin 10.06 and 10.07 (Garcia-Lara et al. 2010; Castro-Alvarez et al. 2015). Fine-mapping of the regions 6.05 and 10.06, and perhaps 3.01, would help to better understand maize defense mechanisms implicated in maize resistance against weevil attack.

As the role of cell-wall fortification as mechanism of defense against insect attack is assumed, genes involved in cell-wall biochemistry and located close to QTL for resistance have been proposed as candidate genes: cellulose synthase genes *cesa3*, *cesa2, cesa4*, *cesa8, cesa11* in bins 2.07, 2.09, 3.07, 6.05, and 7.01, respectively; and peroxidase genes *px1*, *px13* in bins 2.05, 2.07, 2.09, and 5.03, respectively (Hazen et al. 2003; Holland et al. 2000; Garcia-Lara et al. 2009). Other candidate genes suggested are: a gene related to pectin methyl-esterase and pectin content in bin 10.04; genes involved in endosperm softness such as *sen1, sen3,* and *sen5* located in bin 3.01, 1.09, and 2.05, respectively; an specific embryo gene in bin 4.08; the floury grain gene *fl*-N1163* in bin 8.05; and the gene *ZmPrx35* (located between 86,284,199 and 86,285,519 bp on chromosome 10 in the B73 RefGen_V2_sequence) that encodes the peroxidase B6T173 which accounts for most of the peroxidase activity in kernels of the maize resistant population p84C3, a population that underwent three cycles of selection for incremented insect resistance (Castro-Alvarez et al. 2015; Lopez-Castillo et al. 2015). *ZmPrx35* is likely involved in cell-wall strengthening by oxidative coupling of feruloyl polysaccharides.

11.6 Relationships Between Maize Genomics for Resistance to Insects and Agronomical Performance

Adaptation of tropical crops such as maize to temperate climates could decrease defenses against insect attack for two reasons: Insect pressure is reduced in the new environments compared to the original tropical environment, and shorter lifetime of plants in temperate regions reduces the probability of insect attack (Feeny 1976; Descombes et al. 2017). In other words, life span is central in the important trade-off between plant growth rate and plant protection (defenses). Species with longer cycles tend to invest significant resources in protection and, partly as a consequence, grow more slowly. Therefore, the significant genetic correlation between maize earliness and susceptibility to herbivorous attack found in some studies was not surprising (Bohn et al. 2000; Krakowsky et al. 2004; Foiada et al. 2015), mostly when genes for earliness and resistance could be tightly linked as it has been proved for another tropical crop, rice (Yara et al. 2010). Co-localization of QTL for earliness and susceptibility to insect attack has been also reported in maize and fine-mapping of those genomic regions involved in both traits will

likely provide molecular markers to unlink genes for each trait (Krakowsky et al. 2004; Ordas et al. 2010; Foiada et al. 2015; Samayoa et al. 2015a).

More detailed elucidations deserve the possible negative genetic relationship between yield and resistance to insects. Experimental data provide strong correlative support for the notion that trade-offs may exist between selection for growth and yield versus defense against herbivores (Rosenthal and Dirzo 1997). Johnson and Dowd (2004) showed that the overexpression of a single gene regulator activated a defensive pathway sufficient to increase resistance to insects but this activation conferred a cost in plant productivity. In maize, genetic linkage between yield and resistance has been often observed (Klenke et al. 1986; Kreps et al. 1998; Butron et al. 2012). However, genes with pleiotropic effects on both traits or linked genes in repulsion phase have not been yet identified, although QTL with contrasting additive and/or dominance effects on yield and resistance to stalk tunneling by corn borers have been co-localized in several studies (Bohn et al. 1996, 2000; Papst et al. 2001; Ordas et al. 2010). Decreases in growth and photosynthesis in response to insects and pathogens are more likely the result of programmed down-regulation triggered by changes in plant hormone levels rather than the result of metabolic drain toward secondary metabolism involved in plant defense (Foyer et al. 2007; Eichmann and Schafer 2015). These authors suggest that disassembling the underlying hormone signaling networks of growth under insect attack will help to unlink yield and resistance (Eichmann and Schafer 2015). In maize, fine-mapping of genomic regions such as the region 8.04–8.05 could shed light on the relationship between genomics of yield and resistance. In this case, QTL with additive and opposite effects on yield under infestation and resistance co-localized with a QTL for yield mid-parent heterosis (Ordas et al. 2010; Schon et al. 2010; Jímenez-Galindo et al. 2017; Samayoa et al. 2017).

Acknowledgements Investigations done by our research group and reported in the current review have been funded by several projects of the National Plan for Research and Development of Spain, the project AGL2015-67313-C2-1-R being the most recent and funded in part by the European Regional Development Fund. B Ordás and R Santiago acknowledge their grants from the "Ramón y Cajal" program. B Ordás acknowledges CC Schön for providing the unpublished locations of QTL in Foiada et al. 2015.

References

Ahmad S, Veyrat N, Gordon-Weeks R et al (2011) Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. Plant Physiol 157:317–327. https://doi.org/10.1104/pp.111.180224

Ahmed S, Faridi U, Shahid M, Nisar MS (2013) Effect of phenolics in maize grains of different varieties on biology of Angoumois grain moth, *Sitotroga cerealella* (Oliv.) (Lepidoptera: Gelechiidae). Cereal Res Commun 41:636–646

Ajala SO, Kling JG, Schulthess F, et al (2001) Progress in breeding for resistance to maize stem borers *Sesamia calamistis* and *Eldana saccharina* in west and central Africa. In: Seventh Eastern and Southern African regional maize conference, 11–15 Feb 2001, pp 49–54

Arnason JT, Baum B, Gale J et al (1994) Variation in resistance of Mexican landraces of maize to maize weevil, *Sitophilus zeamais*, in relation to taxonomic and biochemical parameters. Euphytica 74:227–236. https://doi.org/10.1007/BF00040405

Arnason JT, Gale J, Conilh de Beyssac B et al (1992) Role of phenolics in resistance of maize grain to the stored grain insects, *Prostephanus truncatus* (Horn) and *Sitophilus zeamais* (Motsch.). J Stored Prod Res 28:119–126. https://doi.org/10.1016/0022-474X(92)90019-M

Barros-rios J, Malvar RA, Jung HG, Santiago R (2011) Phytochemistry cell wall composition as a maize defense mechanism against corn borers. Phytochemistry 72:365–371. https://doi.org/10.1016/j.phytochem.2011.01.004

Barrow MR (1987) The effect of first generation maize stalkborer *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae), on yield of differen maize genotypes. J Entomol Soc South Afr 50:291–298

Barrow MR (1985) The effect of different maize genotypes on the maize stalk-borer, *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae), feeding in whorl tissue. J Entomol Soc South Afr 48:113–119

Barry D, Antonio AQ, Darrah LL (1995) Registration of Mo45, Mo46, and Mo47 maize germplasm lines with resistance to European corn borer. Crop Sci 35:1232–1233

Bergvinson D, Garcia-Lara S (2004) Genetic approaches to reducing losses of stored grain to insects and diseases. Curr Opin Plant Biol 7:480–485. https://doi.org/10.1016/j.pbi.2004.05.001

Bergvinson DJ, Hamilton RI, Arnason JT (1995) Leaf profile of maize resistance factors to European corn borer, *Ostrinia nubilalis*. J Chem Ecol 21:343–354. https://doi.org/10.1007/BF02036722

Betsiashvili M, Ahern KR, Jander G (2015) Additive effects of two quantitative trait loci that confer *Rhopalosiphum maidis* (corn leaf aphid) resistance in maize inbred line Mo17. J Exp Bot 66:571–578. https://doi.org/10.1093/jxb/eru379

Bing JW, Guthrie WD (1991) Generation mean analysis for resistance in maize to the corn leaf aphid (Homoptera: Aphididae). J Econ Entomol 84:1080–1082. https://doi.org/10.1093/jee/84.3.1080

Bing JW, Guthrie WD, Dicke FF (1992) Genetics of resistance in maize to the corn leaf aphid (Homoptera: Aphididae). J Econ Entomol 85:1476–1479. https://doi.org/10.1093/jee/85.4.1476

Bohn M, Groh S, Khairallah MM et al (2001) Re-evaluation of the prospects of marker-assisted selection for improving insect resistance against *Diatraea* spp. in tropical maize by cross validation and independent validation. Theor Appl Genet 103:1059–1067

Bohn M, Khairallah M-M, Jiang C et al (1997) QTL mapping in tropical maize: II. Comparison of genomic regions for resistance to *Diatraea* sp. Crop Sci 37:1892–1902

Bohn M, Khairallah MM, González-de-León D et al (1996) QTL mapping in tropical maize: I. Genomic regions affecting leaf feeding resistance to sugarcane borer and other traits. Crop Sci 36:1352–1361. https://doi.org/10.2135/cropsci1996.0011183X003600050045x

Bohn MO, Marroquin JJ, Flint-Garcia S, Dashiell K, Willmot DB, Hibbard BE (2018) quantitative trait loci mapping of western corn rootworm (Coleoptera: Chrysomelidae) host plant resistance in two populations of doubled haploid lines in maize (Zea mays L.). J Econ Entomol 111:435–444

Bohn M, Schulz B, Kreps R et al (2000) QTL mapping for resistance against the European corn borer (*Ostrinia nubilalis* H.) in early maturing European dent germplasm. Theor Appl Genet 101:907–917. https://doi.org/10.1007/s001220051561

Brooks TD, Bushman BS, Williams WP et al (2007) Genetic basis of resistance to fall armyworm (Lepidoptera: Noctuidae) and southwestern corn borer (Lepidoptera: Crambidae) leaf-feeding damage in maize. J Econ Entomol 100:1470–1475. https://doi.org/10.1603/0022-0493(2007)100%5b1470:gbortf%5d2.0.co;2

Brooks TD, Willcox MC, Williams WP, Buckley PM (2005) Quantitative trait loci conferring resistance to fall armyworm and southwestern corn borer leaf feeding damage. Crop Sci 45:2430–2434. https://doi.org/10.2135/cropsci2004.0656

Bushman BS, Snook ME, Gerke JP et al (2002) Two loci exert major effects on chlorogenic acid synthesis in maize silks. Crop Sci 42:1669–1678

Butrón A, Chen YC, Rottinghaus GE, McMullen MD (2010) Genetic variation at *bx1* controls DIMBOA content in maize. Theor Appl Genet 120:721–734. https://doi.org/10.1007/s00122-009-1192-1

Butron A, Li RG, Guo BZ et al (2001) Molecular markers to increase corn earworm resistance in a maize population. Maydica 46:117–124

Butrón A, Malvar RA, Velasco P et al (1998) Combining abilities and reciprocal effects for maize ear resistance to pink stem borer. Maydica 43:117–122

Butron A, Malvar RA, Cartea ME et al (1999a) Resistance of maize inbreds to pink stem borer. Crop Sci 39:102–107

Butrón A, Malvar RA, Velasco P et al (1999b) Combining abilities for maize stem antibiosis, yield loss, and yield under infestation and non infestation with pink stem borer. Crop Sci 39:691–696

Butron A, Romay MC, Ordas A et al (2008) Genetic and environmental factors reducing the incidence of the storage pest *Sitotroga cerealella* in maize. Entomol Exp Appl 128:421–428. https://doi.org/10.1111/j.1570-7458.2008.00733.x

Butron A, Romay MC, Pena-Asin J et al (2012) Genetic relationship between maize resistance to corn borer attack and yield. Crop Sci 52:1176–1180. https://doi.org/10.2135/cropsci2011.11.0584

Byrne PF, McMullen MD, Snook ME et al (1996) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. Proc Natl Acad Sci USA 93:8820–8825. https://doi.org/10.1073/pnas.93.17.8820

Byrne PF, McMullen MD, Wiseman BR et al (1998) Maize silk maysin concentration and corn earworm antibiosis: QTLs and genetic mechanisms. Crop Sci 38:461–471

Byrne PF, McMullen MD, Wiseman BR et al (1997) Identification of maize chromosome regions associated with antibiosis to corn earworm (Lepidoptera: Noctuidae) larvae. J Econ Entomol 90:1039–1045

Cardinal AJ, Lee M (2005) Genetic relationships between resistance to stalk-tunneling by the European corn borer and cell-wall components in maize population B73 x B52. Theor Appl Genet 111:1–7. https://doi.org/10.1007/s00122-004-1831-5

Cardinal AJ, Lee M, Guthrie WD et al (2006) Mapping of factors for resistance to leaf blade feeding by European corn borer (*Ostrinia nubilalis*) in maize. Maydica 51:93–102

Cardinal AJ, Lee M, Sharopova N et al (2001) Tunneling by the European corn borer in maize. Crop Sci 41:835–845

Carena MJ, Glogoza P (2004) Resistance of maize to the corn leaf aphid: a review. Maydica 49:241–254

Carriere Y, Fabrick JA, Tabashnik BE (2016) Can pyramids and seed mixtures delay resistance to *Bt*

crops? Trends Biotechnol 34:291–302. https://doi.org/10.1016/j.tibtech.2015.12.011

Cartea ME, Malvar RA, Vales MI et al (2001) Inheritance of resistance to ear damage caused by *Sesamia nonagrioides* (Lepidoptera: Noctuidae) in maize. J Econ Entomol 94:277–283

Castro-Alvarez FF, William M, Bergvinson DJ, Garcia-Lara S (2015) Genetic mapping of QTL for maize weevil resistance in a RIL population of tropical maize. Theor Appl Genet 128:411–419. https://doi.org/10.1007/s00122-014-2440-6

Chang S, Brewbaker J (1976) The genetics of resistance to the corn leaf aphid, *Rhopalosiphum maidis* (Fitch). Maize Genet Coop Newsl 31–32

Ching A, Dhugga KS, Appenzeller L et al (2006) *Brittle stalk 2* encodes a putative glycosylphosphatidylinositol-anchored protein that affects mechanical strength of maize tissues by altering the composition and structure of secondary cell walls. Planta 224:1174–1184. https://doi.org/10.1007/s00425-006-0299-8

Chuang WP, Herde M, Ray S et al (2014) Caterpillar attack triggers accumulation of the toxic maize protein RIP2. New Phytol 201:928–939. https://doi.org/10.1111/nph.12581

Cortes-Cruz M, Snook M, McMmullen MD (2003) The genetic basis of C-glycosyl flavone B-ring modification in maize (*Zea mays* L.) silks. Genome 46:182–194. https://doi.org/10.1139/g02-119

Cui ZH, Luo JH, Qi CY, et al (2016) Genome-wide association study (GWAS) reveals the genetic architecture of four husk traits in maize. BMC Genomics 946. https://doi.org/10.1186/s12864-016-3229-6

Dari S, Pixley KV, Setimela P (2010) Resistance of early generation maize inbred lines and their hybrids to maize weevil [*Sitophilus zeamais* (Motschulsky)]. Crop Sci 50:1310–1317. https://doi.org/10.2135/cropsci2009.10.0621

Descombes P, Marchon J, Pradervand JN et al (2017) Community-level plant palatability increases with elevation as insect herbivore abundance declines. J Ecol 105:142–151. https://doi.org/10.1111/1365-2745.12664

Dhliwayo T, Pixley KV (2003) Divergent selection for resistance to maize weevil in six maize populations. Crop Sci 43:2043–2049

Dhliwayo T, Pixley KV, Kazembe V (2005) Combining ability for resistance to maize weevil among 14 southern African maize inbred lines. Crop Sci 45:662–667

Dicke FF (1977) The most important corn insects. In: Sprague GF (ed) Corn and corn improvement. American Society of Agronomy, Inc., Madison, pp 501–590

Dobie P (1974) Laboratory assessment of inherent susceptibility of maize varieties to post-harvest infestation by *Sitophilus zeamais* Motsch (Coleoptera, Curculionidae). J Stored Prod Res 10:183–197. https://doi.org/10.1016/0022-474X(74)90006-X

Eichmann R, Schafer P (2015) Growth versus immunity—a redirection of the cell cycle? Curr Opin Plant Biol 26:106–112. https://doi.org/10.1016/j.pbi.2015.06.006

Feeny P (1976) Plant apparency and chemical defense. In: Wallace JW, Mansell RL (eds) Biochemical interaction between plants and insects. Springer US, Boston, MA, pp 1–40

Fescemyer HW, Sandoya GV, Gill TA et al (2013) Maize toxin degrades peritrophic matrix proteins and stimulates compensatory transcriptome responses in fall armyworm midgut. Insect Biochem Mol Biol 43:280–291. https://doi.org/10.1016/j.ibmb.2012.12.008

Flint-Garcia SA, Darrah LL, McMullen MD, Hibbard BE (2003) Phenotypic versus marker-assisted selection for stalk strength and second-generation European corn borer resistance in maize. Theor Appl Genet 107:1331–1336. https://doi.org/10.1007/s00122-003-1387-9

Flint-Garcia SA, Thuillet AC, Yu JM et al (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Foiada F, Westermeier P, Kessel B 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 128:875–891. https://doi.org/10.1007/s00122-015-2477-1

Foyer CH, Noctor G, van Emden HF (2007) An evaluation of the costs of making specific secondary metabolites: does the yield penalty incurred by host plant resistance to insects result from competition for resources? Int J Pest Manag 53:175–182. https://doi.org/10.1080/09670870701469146

Frey M, Schullehner K, Dick R et al (2009) Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. Phytochemistry 70:1645–1651. https://doi.org/10.1016/j.phytochem.2009.05.012

Garcia-Lara S, Bergvinson DJ (2014) Phytochemical and nutraceutical changes during recurrent selection for storage pest resistance in tropical maize. Crop Sci 54:2423–2432. https://doi.org/10.2135/cropsci2014.03.0223

Garcia-Lara S, Burt AJ, Arnason JT, Bergvinson DJ (2010) QTL mapping of tropical maize grain components associated with maize weevil resistance. Crop Sci 50:815–825. https://doi.org/10.2135/cropsci2009.07.0415

Garcia-Lara S, Khairallah MM, Vargas M, Bergvinson DJ (2009) Mapping of QTL associated with maize weevil resistance in tropical maize. Crop Sci 49:139–149. https://doi.org/10.2135/cropsci2007.06.0326

Gernert WB (1917) Aphis immunity of teosinte-corn hybrids. Science 46:390–392. https://doi.org/10.1126/science.46.1190.390

Gierl A, Frey M (2001) Evolution of benzoxazinone biosynthesis and indole production in maize. Planta 213:493–498. https://doi.org/10.1007/s004250100594

Glauser G, Marti G, Villard N et al (2011) Induction and detoxification of maize 1,4-benzoxazin-3-ones by insect herbivores. Plant J 68:901–911. https://doi.org/10.1111/j.1365-313X.2011.04740.x

Groh S, Gonzalez-de-Leon D, Khairallah MM et al (1998a) QTL mapping in tropical maize: III. Genomic regions for resistance to *Diatraea* spp and associated traits in two RIL populations. Crop Sci 38:1062–1072

Groh S, Khairallah MM, Gonzales-de-Leon D et al (1998b) Comparison of QTL mapped in RIL and their testcross progenies of tropical maize for insect resistance and agronomic traits. Plant Breed 117:193–202

Guo BZ, Widstrom NW, Wiseman BR et al (1999) Comparison of silk maysin, antibiosis to corn earworm larvae (Lepidoptera: Noctuidae), and silk browning in crosses of dent x sweet corn. J Econ Entomol 92:746–753

Guo BZ, Zhang ZJ, Li RG et al (2001) Restriction fragment length polymorphism markers associated with silk maysin, antibiosis to corn earworm (Lepidoptera: Noctuidae) larvae, in a dent and sweet corn cross. J Econ Entomol 94:564–571. https://doi.org/10.1603/0022-0493-94.2.564

Handrick V, Robert CAM, Ahern KR, et al (2016) Biosynthesis of 8-O-methylated benzoxazinoid defense compounds in maize. Plant Cell 28: tpc.00065.2016. https://doi.org/10.1105/tpc.16.00065

Hardwick DF (1965) The corn earworm complex. Mem Entomol Soc Canada 97:5–247

Hawk JA (1985) Registration of DE811 germplasm line of maize. Crop Sci 25:716

Hazen SP, Hawley RM, Davis GL et al (2003) Quantitative trait loci and comparative genomics of cereal cell wall composition. Plant Physiol 132:263–271. https://doi.org/10.1104/pp.103.020016

Hedin PA, Williams WP, Buckley PM, Davis FM (1993) Arrestant responses of southwestern corn borer larvae to free amino acids: structure-activity relationships. J Chem Ecol 19:301–311

Hessel D (2014) Deciphering the genetic architecture of native resistance and tolerance to western corn rootworm larval feeding. Graduate Theses and Dissertations

Holland N, Holland D, Helentjaris T et al (2000) A comparative analysis of the plant cellulose synthase (*CesA*) gene family. Plant Physiol 123:1313–1323. https://doi.org/10.1104/pp.123.4.1313

Jampatong C, McMullen MD, Barry BD et al (2002) Quantitative trait loci for first-and second-generation European corn borer resistance derived from the maize inbred Mo47. Crop Sci 42:584–593

Jiménez-Galindo JC (2017) Regiones genómicas asociadas a la resistencia a insectos y al rendimiento en maíz. PhD Thesis, Universidad de Santiago de Compostela, Spain

Jímenez-Galindo JC, Ordás B, Butron A et al (2017) QTL mapping for yield and resistance against Mediterranean corn borer in maize. Front Plant Sci. https://doi.org/10.3389/fpls.2017.00698

Johnson ET, Dowd PF (2004) Differentially enhanced insect resistance, at a cost, in *Arabidopsis thaliana* constitutively expressing a transcription factor of defensive metabolites. J Agric Food Chem 52:5135–5138. https://doi.org/10.1021/jf0308049

Khairallah MM, Bohn M, Jiang C et al (1998) Molecular mapping of QTL for southwestern corn borer resistance, plant height and flowering in tropical maize. Plant Breed 117:309–318

Klenke JR, Russell WA, Guthrie WD (1986) Grain-yield reduction caused by 2nd generation European corn-borer in BS9 corn synthetic. Crop Sci 26:859–863

Krakowsky MD, Lee M, Holland JB (2007) Genotypic correlation and multivariate QTL analyses for cell wall components and resistance to stalk tunneling by the European corn borer in maize. Crop Sci 47:485–490. https://doi.org/10.2135/cropsci2006.05.0283

Krakowsky MD, Lee M, Woodman-Clikeman WL et al (2004) QTL mapping of resistance to stalk tunneling by the European corn borer in RILs of maize population B73 x De811. Crop Sci 44:274–282

Kreps RC, Gumber RK, Schulz B et al (1998) Genetic variation in testcrosses of European maize inbreds for resistance to the European corn borer and relations to line per se performance. Plant Breed 117:319–327. https://doi.org/10.1111/j.1439-0523.1998.tb01948.x

Kumar F (1994a) Mechanisms of resistance in maize to southwestern corn borer and sugarcane borer. In: Mihm JA (ed) Insect resistant maize, proceeding of an international symposium held at the CIMMYT. CIMMYT, Mexico, p 304

Kumar F (1994b) An overview of research on mechanisms of resistance in maize to spotted stem borer. In: Mihm JA (ed) Insect resistant maize, proceeding of an international symposium held at the CIMMYT. Mexico, p 304

Kumar H (2002) Resistance in maize to the larger grain borer, *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae). J Stored Prod Res 38:267–280. https://doi.org/10.1016/S0022-474X(01)00030-3

Kumar H, Mihm JA (1996) Resistance in maize hybrids and inbreds to first-generation southwestern corn borer, *Diatraea grandiosella* (Dyar) and sugarcane borer, *Diatraea saccharalis* Fabricius. Crop Prot 15:311–317

Lee EA, Byrne PF, McMullen MD et al (1998) Genetic mechanisms underlying apimaysin and maysin synthesis and corn earworm antibiosis in maize (*Zea mays* L.). Genetics 149:1997–2006

Li X, He KL, Wang ZY, Bai SX (2010) Quantitative trait loci for Asian corn borer resistance in maize population Mc37 x Zi330. Agric Sci China 9:77–84. https://doi.org/10.1016/s1671-2927(09)60070-5

Liu KJ, Goodman M, Muse S et al (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Long BJ, Dunn GM, Bowman JS, Routley DG (1977) Relationship of hydroxamic acid content in corn and resistance to the corn leaf aphid. Crop Sci 17:55. https://doi.org/10.2135/cropsci1977.0011183X001700010016x

Lopez-Castillo LM, Lopez-Arciniega JI, Guerrero-Rangel A et al (2015) Identification of B6T173 (ZmPrx35) as

the prevailing peroxidase in highly insect-resistant maize (*Zea mays*, p84C3) kernels by activity-directed purification. Front Plant Sci. https://doi.org/10.3389/fpls.2015.00670

Lu X (1999) Identification and mapping of quantitative trait loci conferring disease and insect resistance in maize. PhD thesis, University of Hawaii

Lu X, Brewbaker J (1999) Genetics of resistance in maize to the corn leaf aphid (Homoptera: Aphididae). Maize Genet Coop Newsl 73:36–37

Makowska B, Bakera B, Rakoczy-Trojanowska M (2015) The genetic background of benzoxazinoid biosynthesis in cereals. Acta Physiol Plant 37:176. https://doi.org/10.1007/s11738-015-1927-3

Malvar RA, Cartea ME, Revilla P et al (1993) Sources of resistance to pink stem borer and European corn borer in maize. Maydica 38:313–319

McMullen MD, Byrne PF, Snook ME et al (1998) Quantitative trait loci and metabolic pathways. Proc Natl Acad Sci USA 95:1996–2000. https://doi.org/10.1073/pnas.95.5.1996

McMullen MD, Frey M, Degenhardt J (2009) Genetics and biochemistry of insect resistance in maize. In: Bennetzen JL, Hake SC (eds) Handbook of maize: its biology. Springer, pp 271–289

McMullen MD, Kross H, Snook ME et al (2004) Salmon silk genes contribute to the elucidation of the flavone pathway in maize (*Zea mays* L.). J Hered 95:225–233. https://doi.org/10.1093/jhered/esh042

McMullen MD, Snook M, Lee EA et al (2001) The biological basis of epistasis between quantitative trait loci for flavone and 3-deoxyanthocyanin synthesis in maize (*Zea mays* L.). Genome 44:667–676. https://doi.org/10.1139/gen-44-4-667

Meihls LN, Handrick V, Glauser G et al (2013) Natural variation in maize aphid resistance is associated with 2,4-Dihydroxy-7-Methoxy-1,4-Benzoxazin-3-One glucoside methyltransferase activity. Plant Cell 25:1–16. https://doi.org/10.1105/tpc.113.112409

Melchinger AE, Kreps R, Spath R et al (1998) Evaluation of early-maturing European maize inbreds for resistance to the European corn borer. Euphytica 99:115–125. https://doi.org/10.1023/a:1018305316853

Meyer JDF, Snook ME, Houchins KE et al (2007) Quantitative trait loci for maysin synthesis in maize (*Zea mays* L.) lines selected for high silk maysin content. Theor Appl Genet 115:119–128. https://doi.org/10.1007/s00122-007-0548-7

Midega CAO, Murage AW, Pittchar JO, Khan ZR (2016) Managing storage pests of maize: farmers' knowledge, perceptions and practices in western Kenya. Crop Prot 90:142–149. https://doi.org/10.1016/j.cropro.2016.08.033

Mihm JA (1985) Breeding for host plant resistance to maize stem-borers. Int J Trop Insect Sci 6:369–377. https://doi.org/10.1017/S1742758400004653

Mohan S, Ma PWK, Williams WP, Luthe DS (2008) A naturally occurring plant cysteine protease possesses remarkable toxicity against insect pests and synergizes *Bacillus thuringiensis* toxin. PLoS One. https://doi.org/10.1371/journal.pone.0001786

Moose SP, Sisco PH (1996) Glossy15, an APETALA2-like gene from maize that regulates leaf epidermal cell identity. Genes Dev 10:3018–3027. https://doi.org/10.1101/gad.10.23.3018

Mwololo JK, Mugo SN, Tefera T et al (2012) Resistance of tropical maize genotypes to the larger grain borer. J Pest Sci (2004) 85:267–275. https://doi.org/10.1007/s10340-012-0427-0

Niemeyer HM (2009) Hydroxamic acids derived from 2-hydroxy-2H-1,4-benzoxazin-3(4H)-one: key defense chemicals of cereals. J Agric Food Chem 57:1677–1696. https://doi.org/10.1021/jf8034034

Nwosu LC (2016) Chemical bases for maize grain resistance to infestation and damage by the maize weevil, *Sitophilus zeamais* Motschulsky. J Stored Prod Res 69:41–50. https://doi.org/10.1016/j.jspr.2016.06.001

Oikawa A, Ishihara A, Tanaka C et al (2004) Accumulation of HDMBOA-Glc is induced by biotic stresses prior to the release of MBOA in maize leaves. Phytochemistry 65:2995–3001. https://doi.org/10.1016/j.phytochem.2004.09.006

Ordas B, Malvar RA, Santiago R et al (2009) Mapping of QTL for resistance to the Mediterranean corn borer attack using the intermated B73 x Mo17 (IBM) population of maize. Theor Appl Genet 17:1451–1459. https://doi.org/10.1007/s00122-009-1147-6

Ordas B, Malvar RA, Santiago R, Butron A (2010) QTL mapping for Mediterranean corn borer resistance in European flint germplasm using recombinant inbred lines. BMC Genomics 174. https://doi.org/10.1186/1471-2164-11-174

Orsini E, Krchov LM, Uphaus J, Melchinger AE (2012) Mapping of QTL for resistance to first and second generation of European corn borer using an integrated SNP and SSR linkage map. Euphytica 183:197–206. https://doi.org/10.1007/s10681-011-0441-y

Overholt WA, Ogedah K, Lammers PM (1994) Dstribution and sampling of *Chilo partellus* (Lepidoptera, Pyralidae) in maize and sorghum on the Kenya coast. Bull Entomol Res 84:367–378

Papst C, Bohn M, Utz HF et al (2004) QTL mapping for European corn borer resistance (*Ostrinia nubilalis* Hb.), agronomic and forage quality traits of testcross progenies in early-maturing European maize (*Zea mays* L.) germplasm. Theor Appl Genet 108:1545–1554. https://doi.org/10.1007/s00122-003-1579-3

Papst C, McKhinger AE, Eder J et al (2001) QTL mapping for resistance to European corn borer (*Ostrinia nubilalis* HB.) in early maturing European dent maize (*Zea mays* L.) germplasm and comparison of genomic regions for resistance across two populations of F-3 families. Maydica 46:195–205

Pechan T, Ye L, Chang Y et al (2000) A unique 33-kD cysteine proteinase accumulates in response to larval feeding in maize genotypes resistant to fall armyworm and other Lepidoptera. Plant Cell 12:1031–1040

Robert CAM, Ferrieri RA, Schirmer S et al (2014) Induced carbon reallocation and compensatory growth as root herbivore tolerance mechanisms. Plant Cell Environ 37:2613–2622. https://doi.org/10.1111/pce.12359

Rosenthal JP, Dirzo R (1997) Effects of life history, domestication and agronomic selection on plant defence against insects: evidence from maizes and wild relatives. Evol Ecol 11:337–355. https://doi.org/10.1023/a:1018420504439

Samayoa LF, Butron A, Malvar RA (2014) QTL mapping for maize resistance and yield under infestation with *Sesamia nonagrioides*. Mol Breed 34:1331. https://doi.org/10.1007/s11032-014-0119-y

Samayoa LF, Malvar RA, Butrón A (2017) QTL for maize midparent heterosis in the heterotic pattern American dent x European flint under corn borer pressure. Front Plant Sci. https://doi.org/10.3389/fpls.2017.00573

Samayoa LF, Malvar RA, McMullen MD, Butron A (2015a) Identification of QTL for resistance to Mediterranean corn borer in a maize tropical line to improve temperate germplasm. BMC Plant Biol 15:265. https://doi.org/10.1186/s12870-015-0652-9

Samayoa LF, Malvar RA, Olukolu BA et al (2015b) Genome-wide association study reveals a set of genes associated with resistance to the Mediterranean corn borer (Sesamia nonagrioides L.) in a maize diversity panel. BMC Plant Biol 15:35. https://doi.org/10.1186/s12870-014-0403-3

Sandoya G, Butron A, Alvarez A et al (2008) Direct response of a maize synthetic to recurrent selection for resistance to stem borers. Crop Sci 48:113–118. https://doi.org/10.2135/cropsci2007.02.0084

Santiago R, Barros-Rios J, Alvarez A, Malvar RA (2016a) Agronomic performance of maize populations divergently selected for diferulate cross-linkage. J Agric Sci 154:1270–1279. https://doi.org/10.1017/S0021859615001161

Santiago R, Malvar RA, Barros-rios J et al (2016b) Hydroxycinnamate synthesis and asssociation with Mediterranean corn borer resistance. J Agric Food Chem 64:539–551. https://doi.org/10.1021/acs.jafc.5b04862

Schon CC, Dhillon BS, Utz HF, Melchinger AE (2010) High congruency of QTL positions for heterosis of grain yield in three crosses of maize. Theor Appl Genet 120:321–332. https://doi.org/10.1007/s00122-009-1209-9

Schulz B, Kreps R, Klein D et al (1997) Genetic variation among European maize inbreds for resistance to the European corn borer and relation to agronomic traits. Plant Breed 116:415–422. https://doi.org/10.1111/j.1439-0523.1997.tb01024.x

Scott G, Davis F (1978) Second-brood damage by southwestern corn-borer in a corn diallel cross. Crop Sci 18:335–336

Sekhon SS, Kanta U (1994) Mechanisms and bases of resistance in maize to spotted stem borer. In: Mihm JA (ed) Insect Resistant Maize, proceeding of an international symposium held at the CIMMYT. CIMMYT, Mexico, p 304

Sharma M, Chai CL, Morohashi K, et al (2012) Expression of flavonoid 3'-hydroxylase is controlled by P1, the regulator of 3-deoxyflavonoid biosynthesis in maize. BMC Plant Biol 196. https://doi.org/10.1186/1471-2229-12-196

Snook ME, Wiseman BR, Widstrom NW, Wilson RL (1997) Chemicals associated with maize resistance to corn earworm and fall armyworm. In: Proceedings of an International Symposium, Mexico, DF (Mexico), 27 Nov–3 Dec 1994. CIMMYT

So Y (2003) Corn leaf aphid and polysora rust resistance in tropical maize. Master's thesis, University of Hawaii

So YS, Ji HC, Brewbaker JL (2010) Resistance to corn leaf aphid (*Rhopalosiphum maidis* Fitch) in tropical corn (*Zea mays* L.). Euphytica 172:373–381. https://doi.org/10.1007/s10681-009-0044-z

Szalma SJ, Buckler ES, Snook ME, McMullen MD (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor Appl Genet 110:1324–1333. https://doi.org/10.1007/s00122-005-1973-0

Szalma SJ, Snook ME, Bushman BS et al (2002) Duplicate loci as QTL: the role of chalcone synthase loci in flavone and phenylpropanoid biosynthesis in maize. Crop Sci 42:1679–1687

Tabashnik BE, Roush RT, Earle ED, Shelton AM (2000) Resistance to Bt toxins. Science 287:42

Tefera T, Mugo S, Beyene Y (2016) Developing and deploying insect resistant maize varieties to reduce pre- and post-harvest food losses in Africa. Food Secur 8:211–220. https://doi.org/10.1007/s12571-015-0537-7

Thome CR, Smith ME, Mimh J (1992) Leaf feeding resistance to multiple insect species in a maize diallel. Crop Sci 32:1460–1463. https://doi.org/10.2135/cropsci1992.0011183X003200060030x

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 95:1005–1011. https://doi.org/10.1007/s001220050654

Tzin V, Lindsay PL, Christensen SA et al (2015) Genetic mapping shows intraspecific variation and transgressive segregation for caterpillar-induced aphid resistance in maize. Mol Ecol 24:5739–5750. https://doi.org/10.1111/mec.13418

van Rensburg JBJ, Malan C (1990) Resistance of maize genotypes to the maize stalk borer, *Busseola fusca* (Fuller) (Lepidoptera: Noctuidae). J Entomol Soc S Afr 53:49–55

van Rensburg JBJ, van den Berg J (1995) New sources of resistance to the stalk borers *Busseola fusca* (Fuller) and *Chilo partellus* Swinhoe in maize. S Afr J Plant Soil 12:91–93. https://doi.org/10.1080/02571862.1995.10634343

Velasco P, Revilla P, Butron A et al (2002) Ear damage of sweet corn inbreds and their hybrids under multiple corn borer infestation. Crop Sci 42:724–729

Waiss ACJ, Chan BG, Elliger CA et al (1979) Maysin, a flavone glycoside from corn silks with antibiotic activity toward corn earworm. J Econ Entomol 72:256–258

Willcox MC, Khairallah MM, Bergvinson D et al (2002) Selection for resistance to southwestern corn borer using marker-assisted and conventional backcrossing. Crop Sci 42:1516–1528. https://doi.org/10.2135/cropsci2002.1516

Williams PW, Davis FM, Buckley PM et al (1998) Factors associated with resistance to fall armyworm (Lepidoptera: Noctuidae) and southwestern corn borer (Lepidoptera: Crambidae) in corn at different vegetative stages. J Econ Entomol 91:1471–1480. https://doi.org/10.1093/jee/91.6.1471

Williams W, Buckley P, Davis F (1989) Combinin ability for resistance in corn to fall armyworm and southwestern corn borer. Crop Sci 29:913–915

Williams WP, Buckley PM, Taylor VN (1983) Southwestern corn-borer growth on callus initiated from corn genotypes with different levels of resistance to plant-damage. Crop Sci 23:1210–1212

Williams WP, Davis FM (2000) Registration of maize germplasms Mp713 and Mp714

Williams WP, Davis FM (1994) Mechanisms and bases of resistance in maize to Southwestern corn borer and fall armyworm. In: Mihm JA (ed) Insect resistant maize, proceeding of an international symposium held at the CIMMYT. Mexico, p 304

Wiseman BR, Snook ME, Isenhour DJ et al (1992) Relationship between growth of corn- earworm and fall armyworm larvae (Lepidoptera, Noctuidae) and maysin concentration in corn silks. J Econ Entomol 85:2473–2477

Wiseman BR, Widstrom NW (1992) Resistance of corn populations to larvae of the corn earworm (Lepidoptera: Noctuidae). J Econ Entomol 85:601–605

Wouters FC, Blanchette B, Gershenzon J, Vassao DG (2016) Plant defense and herbivore counter-defense: benzoxazinoids and insect herbivores. Phytochem Rev 15:1127–1151. https://doi.org/10.1007/s11101-016-9481-1

Yan J, Lipka AE, Schmelz EA et al (2015) Accumulation of 5-hydroxynorvaline in maize (*Zea mays*) leaves is induced by insect feeding and abiotic stress. J Exp Bot 66:593–602. https://doi.org/10.1093/jxb/eru385

Yara A, Cong NP, Matsumura M et al (2010) Development of near-isogenic lines for BPH25(t) and BPH26 (t), which confer resistance to the brown planthopper, *Nilaparvata lugens* (Stal.) in indica rice "ADR52". Breed Sci 60:639–647. https://doi.org/10.1270/jsbbs.60.639

Zhang PF, Wang YB, Zhang JB et al (2003) A maize QTL for silk maysin levels contains duplicated Myb-homologous genes which jointly regulate flavone biosynthesis. Plant Mol Biol 52:1–15. https://doi.org/10.1023/a:1023942819106

Zheng L, McMullen MD, Bauer E et al (2015) Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in *Zea mays*. J Exp Bot 66:3917–3930. https://doi.org/10.1093/jxb/erv192

The Genetics and Genomics of Virus Resistance in Maize

12

Margaret G. Redinbaugh, Thomas Lübberstedt, Pengfei Leng and Mingliang Xu

Abstract

Viruses cause significant diseases on maize worldwide. Intensive agronomic practices, changes in vector distribution, and the introduction of vectors and viruses into new areas can result in emerging disease problems. Because deployment of resistant hybrids and cultivars is considered to be both economically viable and environmentally sustainable, genes and quantitative trait loci for most economically important virus diseases have been identified. Examination of multiple studies indicates the importance of regions of maize chromosomes 2, 3, 6, and 10 in virus resistance. An understanding of the molecular basis of virus resistance in maize is beginning to emerge, and two genes conferring resistance to sugarcane mosaic virus, *Scmv1* and *Scmv2*, have been cloned and characterized. Recent studies provide hints of other pathways and genes critical to virus resistance in maize, but further work is required to determine the roles of these in virus susceptibility and resistance. This research will be facilitated by rapidly advancing technologies for functional analysis of genes in maize.

12.1 Introduction

Viruses cause significant disease in crops worldwide (Gomez et al. 2009; Kang et al. 2005), and they account for the majority of emerging diseases in plants (Anderson et al. 2004). In maize, losses due to virus diseases were estimated at 3% (Oerke and Dehne 2004). Based on estimated production of 875 million tonnes worldwide (Ranum et al. 2014), losses would be about 26 billion tonnes of grain valued at about 4.5 billion USD. Although more than 50 virus species can infect maize (Lapierre and Signoret 2004), only about a dozen of these cause significant disease problems (Stewart et al. 2016; Redinbaugh and Zambrano Mendoza 2014).

In contrast to most fungal and bacterial pathogens, viruses are obligate intracellular pathogens, dependent on the host cell for replication (Hull 2002). Plant virus genomes may

The original version of this chapter was revised: For detailed information please see correction. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-97427-9_22

M. G. Redinbaugh (✉)
USDA-ARS Corn, Soybean and Wheat Quality Research Unit and Department of Plant Pathology, The Ohio State University, Wooster, OH 44691, USA
e-mail: redinbaugh.2@osu.edu

T. Lübberstedt · P. Leng
Department of Agronomy, Iowa State University, Ames, IA 50011, USA

P. Leng · M. Xu
National Maize Improvement Center, China Agricultural University, Beijing, China

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes, https://doi.org/10.1007/978-3-319-97427-9_12

consist of double-stranded or single-stranded RNA or DNA, may have a single genome segment or be multipartite, and generally encode fewer than 20 genes. Viruses generally enter plant cells due to mechanical disruption of the cell wall and membrane resulting from insect feeding, abrasion, or other means of wounding. Most maize-infecting viruses are transmitted by Hemipteran insects, but maize chlorotic mottle virus (MCMV) and wheat mosaic virus (WMoV, the causal agent of High Plains disease) are transmitted by thrips or beetles and mites, respectively (Cabanas et al. 2013; Nault et al. 1978; Stenger et al. 2016). Some viruses can also be transmitted through seed (Albrechtsen 2006), but well-documented rates of seed transmission are low for maize-infecting viruses (Johansen et al. 1994). In general, virus diseases occur when a source of virus and competent vectors occurs together with a susceptible host under suitable environmental conditions. Agronomic approaches to virus disease control include chemical control of insect vector populations, adjusting planting dates to avoid vectors, removal of weedy virus reservoirs and crop rotation. However, strong genetic resistance to most maize-infecting viruses has been identified, providing an economically sound, environmentally sustainable approach for disease control.

For the past 50 years, several viruses have caused, and continue to cause, significant agricultural problems in maize. Viruses in the family *Potyviridae*, primarily maize dwarf mosaic virus (MDMV) and sugarcane mosaic virus (SCMV), cause disease on maize everywhere the crop is grown (Stewart et al. 2016). Maize streak, caused by the geminivirus, maize streak virus (MSV), has been known for more than 100 years across sub-Saharan Africa, where it continues to cause significant food insecurity (Martin and Shepherd 2009). The rhabdovirus maize mosaic virus (MMV) was identified as a pathogen in 1960 (Herold et al. 1960), but the disease caused by the virus has been known for centuries in the tropics and sub-tropics where its planthopper vector is prevalent (Brewbaker 1979; Brewbaker 1981; Lapierre and Signoret 2004). The fijiviruses, maize rough dwarf virus (MRDV), and rice black-streaked dwarf virus (RBSDV), first emerged in Europe in the late 1940s (Lapierre and Signoret 2004). These viruses continue to cause crop losses there and in China, where agronomic practices facilitate large populations of viruliferous vectors. In South America, the related fijivirus, Mal de Rio Cuarto virus (MRCV) (Bonamico et al. 2010; Lapierre and Signoret 2004), also causes problems for farmers and seed producers. Disease caused by all of these viruses is controlled, at least to some extent, with resistant or tolerant maize hybrids and cultivars.

The recent emergence of two virus diseases is currently of concern. The most important is maize lethal necrosis (MLN), which results from the synergistic interaction of maize chlorotic mottle virus (MCMV) with another virus, usually from the family *Potyviridae* (Niblett and Claflin 1978). MLN was first described in the 10–70s in Kansas and Nebraska in the USA, where it caused significant but localized problems. Since 2011, however, MLN has rapidly emerged in sub-Saharan East Africa where it can cause up to 100% losses of maize crops (Mahuku et al. 2015; Wangai et al. 2012). MLN has also recently emerged and spread in China, Taiwan and Ecuador (Deng et al. 2014; Quito-Avila et al. 2016; Xie et al. 2011). Disease emergence has been closely tied to the presence of the MCMV vector, maize thrips (*Frankliniella williamsii* Hood), and to multiple annual maize crops (Cabanas et al. 2013; Mahuku et al. 2015). High Plains disease caused by WMoV was first discovered in the 1990s on maize in the US Midwest (Jensen et al. 1996). WMoV continues to cause important disease in wheat, and seed and sweet corn (Stewart et al. 2016). The disease causes problems for seed companies and maize breeders due to a potential for seed transmission (Jensen et al. 1996) that has led to phytosanitary restrictions to seed movement.

In model systems, we have some understanding of the molecular and genomic interactions among the host plant, viral pathogen, and insect vector that lead to virus susceptibility or resistance, but we are just beginning to define these events in cereal crops like maize.

12.2 Genome Sequencing for Virus Diagnostics and Characterization

Increasingly, next-generation sequencing (NGS) approaches are being used to identify viruses and characterize their populations in plants. Because most plant-infecting viruses have RNA genomes (and viruses with DNA genomes still make RNA transcripts), RNA-Seq approaches have been favored for these analyses. For maize, NGS was used to identify MCMV and SCMV in MLN-affected plants (Adams et al. 2013) and was subsequently used to demonstrate sequence homogeneity and diversity among MCMV and SCMV populations, respectively, in MLN-affected maize (Mahuku et al. 2015). NGS was also used to identify Johnsongrass mosaic virus (JGMV) in samples from Kenya and Uganda, and further experiments demonstrated a role for this virus in causing MLN (Stewart et al. 2017). A new polerovirus, tentatively named maize yellow mosaic virus, was identified in southwestern China using NGS (Chen et al. 2016) and was subsequently found in maize from southeastern China, Ecuador, and sub-Saharan Africa (Bernreiter et al. 2017; Palanga et al. 2017; Wang et al. 2016, Stewart et al. *unpublished results*). Similarly, a genome sequence related to fungal totiviruses was identified in maize (Chen et al. 2016). The clear utility of NGS for defining virus sequences and their diversity in crops indicates that this platform will become increasingly valuable as a diagnostic tool. However, the biological and epidemiological roles of the identified viruses must still be characterized to determine the role(s) of specific viruses identified by NGS in disease.

12.3 The Genetics of Virus Resistance in Maize

With few exceptions, maize inbred lines with strong virus resistance have been identified. In these lines, virus inoculation produces no or few symptoms. Importantly, the virus is excluded from or is found at significantly reduced titer in systemic plant tissues. An important exception to this is MCMV. For this virus, tolerant maize inbred lines developing few or no symptoms after inoculation with MCMV have been identified, but the virus is present at high titer in systemic tissues in these lines (Jones et al. 2018). Resistance has been associated with both dominant genes, such as those for resistance to potyviruses, and quantitative trait loci (QTL) with additive or dominant gene action, such as those for resistance to maize chlorotic dwarf virus (MCDV), maize rayado fino virus (MRCV), or maize mosaic virus (MMV) have been identified (Redinbaugh and Zambrano Mendoza 2014). Here again, MCMV is an exception, with a major QTL that have recessive character having been identified in two populations (Jones et al. 2018). Major QTL generally account for more than 20% of the phenotypic variation for resistance, although this is highly dependent on the population. Minor QTL, accounting for less than 10% of the phenotypic variance, have also been identified for resistance to several viruses.

12.3.1 Genetics of Potyvirus Resistance in Maize

Resistance to viruses in the *Potyviridae* has been investigated in US, European, Chinese, and tropical germplasm (reviewed in Redinbaugh and Pratt 2008; Pratt and Gordon 2006; Liu et al. 2009b). A strong correlation between MDMV and SCMV susceptibility was found among 122 European (Kuntze et al. 1997) and 155 U.S. and tropical (Jones et al. 2007) maize inbreds. Only three European lines (D21, D32, and FAP1360A) displayed complete resistance to SCMV and MDMV. The US line Pa405 and the Caribbean line Oh1VI are both completely resistant to MDMV, SCMV, and wheat streak mosaic virus (WSMV; Louie et al. 1991; Zambrano et al. 2014). Although minor gene resistance to these viruses has been identified in some lines, major loci for resistance have been identified in three genomic regions in all germplasm tested. Resistance to MDMV in Pa405 is conferred by a dominant resistance gene, *Mdm1*,

which mapped to the short arm of chromosome (chr.) 6 (McMullen and Louie 1989). Two or three genes involved in resistance to SCMV were identified in different crosses, and two dominant major resistance genes, *Scmv1* and *Scmv2*, were mapped on the short arm of chr. 6, and near the centromere of chr. 3, respectively (Melchinger et al. 1998). These resistance genes interact epistatically and are simultaneously required for expression of complete resistance to SCMV. *Scmv1* provides resistance at all developmental stages, and *Scmv2* is expressed at later stages of plant development (Xia et al. 1999). Pa405 carries three genes for resistance to WSMV: *Wsm1* on the short arm of chr. 6, *Wsm2* near the centromere of chr. 3, and *Wsm3* on the long arm of chr. 10 (McMullen et al. 1994). Thus, *Mdm1*, *Scmv1*, and *Wsm1* map to the same location, as do *Scmv2* and *Wsm2*.

In four separate studies, near-isogenic lines (NIL) carrying the *Scmv1* and *Scmv2*, or *Wsm1*, *Wsm2,* and *Wsm3* genes in various combinations, were tested for their responses to potyvirus species and isolates (Jones et al. 2011; Lubberstedt et al. 2006; Stewart et al. 2013; Xing et al. 2006) (Table 12.1). Inoculation of isogenic homozygous lines carrying resistance or susceptibility alleles derived from FAP1360A at *Scmv1* or *Scmv2* with the Seehausen isolate of SCMV (SCMV-Gr) and an Israeli isolate of MDMV (MDMV-MD) indicated that single gene was insufficient for resistance to either virus (Xing et al. 2006). The $F7^{RR/RR}$ line carrying both genes was completely resistant to SCMV-Gr, the Ohio isolate of MDMV (MDMV-OH), MDMV-MD, and the Ohio isolate of WSMV (WSMV-OH) (Lubberstedt et al. 2006). However, this line was susceptible to the Ohio SCMV

Table 12.1 Responses of lines carrying resistance loci on chromosomes 6, 3, and 10 to inoculation with potyvirus isolates[a]

			Chr[d]			*Virus isolate*[e]							
S recurrent parent	R source[b]	Line[c]	3	6	10	M-OH	M-It	M-MD	S-OH	S-Gr	J-Tx	Sr	W
F7	FAP1360A	$F7^{RR/RR}$	✓	✓		R[f]	S	R	S	R	–	–	R
		$F7^{RR/SS}$	✓			S	S	S	S	S	–	–	–
		$F7^{SS/RR}$		✓		R	S	S	S	V	–	–	–
		F7 control				S	S	S	S	S	–	–	S
Oh28	Pa405	$Oh28^{RR/RR/SS}$	✓	✓		R	R	–	R	R	R	R	R
		$Oh28^{SS/RR/RR}$		✓	✓	R	–	–	S	–	R	R	R
		$Oh28^{SS/RR/SS}$		✓		R	S	–	S	R	R	R	R
		$Oh28^{RR/SS/SS}$	✓			S	S	–	S	S	S	S	R
		$Oh28^{SS/SS/RR}$			✓	S	–	–	S	–	S	S	R
		Oh28 control				S	S	–	S	S	S	S	S

[a]The results presented are summarized from Jones et al. (2011), Lübberstedt et al. (2006), Stewart et al. (2013), Xing et al. (2006)

[b]The potyvirus resistant inbred line used as donor parent to generate near-isogenic lines. F7 and Oh28 were the potyvirus susceptible lines used as recurrent parents

[c]Near-isogenic lines (NIL) with resistance loci introgressed from the indicated resistance source. The superscripts xx/yy and xx/yy/zz indicate the presence of resistance (R) or susceptible (S) alleles on chr. 3 (x), 6 (y), and 10 (z)

[d]Chromsome; the check marks indicate the presence of resistance loci from chromosome 3, 6, or 10

[e]The virus isolate tested. *M-OH* maize dwarf mosaic virus (MDMV) Ohio isolate; *M-It* MDMV Italian isolate; *M-MD* MDMV Israel isolate; *S-OH* sugarcane mosaic virus (SCMV) Ohio isolate; *S-Gr* SCMV Seehausen isolate; *J-Tx* Johnsongrass mosaic virus Texas isolate; *Sr* sorghum mosaic virus Texas isolate; *W* WSMV Ohio isolate; *Wo* wheat mosaic virus Kansas isolate

[f]*R* resistant; *S* susceptible; *V* variable, expressing resistance at 7 dpi and susceptibility at 14 dpi; – not tested

isolate (SCMV-OH) and an aggressive isolate of MDMV from Italy (MDMV-It). Lines carrying *Scmv1* alone ($F7^{SS/RR}$) provided resistance to MDMV-OH and early resistance to SCMV-Gr, but the *Scmv2* gene alone ($F7^{RR/SS}$) did not provide any resistance. NIL carrying the *Wsm1* gene from Pa405 ($Oh28^{SS/RR/SS}$) were resistant to MDMV-OH and SCMV-Gr, but not MDMV-It or SCMV-OH. However, lines carrying both *Wsm1* and *Wsm2* ($Oh28^{RR/RR/SS}$) were resistant to all potyviruses tested. These results suggest that the Pa405-derived allele on chr. 6 (*Wsm1*) is stronger than the allele from FAP1360A (*Scmv1*). Although the patterns of resistance are similar for FAP1360A- and Pa405-derived isogenic lines, these two inbred lines are only distantly related (Xu et al. 2000).

Although NIL homozygous for *Wsm1* were completely resistant to MDMV-OH, epistatic resistance from *Wsm2* and *Wsm3*, or closely linked genes, was detected in NIL heterozygous for *Wsm1* (Jones et al. 2011). NIL carrying *Wsm1* were resistant to JGMV and sorghum mosaic virus (SrMV), and neither *Wsm2* nor *Wsm3* provided resistance on their own. Any of the three genes, *Wsm1, Wsm2,* or *Wsm3,* provided complete resistance to WSMV (McMullen et al. 1994). Together the results suggest that potyviruses and potyvirus isolates can vary in their virulence against the resistance genes on chr. 3 and 6, and indicate a relative virulence of (SCMV-OH ~ MDMV-It) > (SCMV-Gr ~ MDMV-OH ~ JGMV ~ SrMV) > WSMV. With the significant genomic sequence diversity among these viruses, it is of interest to identify the virus factors that influence virulence to determine whether conserved nucleotide or protein sequences, or conserved three-dimensional structures play roles in virus species and isolate virulence.

12.3.2 Resistance to Other Viruses in Other Families

In contrast to the potyviruses, which are easily mechanically transmitted under greenhouse and field conditions, many of the other important maize-infecting viruses must be transmitted using insect vectors, or more specialized techniques like agro-infiltration (Boulton et al. 1989) or vascular puncture inoculation (Louie 1995). Despite the difficulties associated with assessing phenotypic responses for obligately insect-vectored viruses, the genetics of resistance to at least eight virus diseases caused by potyviruses and eleven other virus species has been characterized (Redinbaugh and Zambrano Mendoza 2014). With our rapidly evolving resources for genotyping maize populations (Elshire et al. 2011; Ganal et al. 2011), genetic characterization of resistance has become limited only by our ability to develop populations and the ability to implement phenotypic analyses. Increasingly, genotyped association mapping populations, including the nested association mapping population, are available to researchers (Flint-Garcia et al. 2005; McMullen et al. 2009; Romay et al. 2013). These populations may prove invaluable for identifying virus resistance loci in maize, if a sufficient proportion of the population carries virus resistance.

Virus resistance loci have been found on nine of the ten maize chromosomes. By estimating the physical positions of markers for virus resistance QTL on the B73 v3 genome, results of previous studies were combined to identify nine clusters of virus resistance loci on chr. 1, 2, 3, 6, 8, and 10 (Table 12.2). While the same QTL for a given virus may have been identified in a number of studies (reviewed in Redinbaugh and Zambrano (2014), the positions of only the most well-defined QTL for each virus were included in Table 12.2.

Five of the resistance QTL are associated with a single virus, and three of these are for resistance to maize streak virus (MSV), which is the only DNA virus currently causing disease problems in maize. These are found in bins 1.06, 2.06, and 3.09 (Nair et al. 2015; Welz et al. 1998). The other two individual QTL are for tolerance to MCMV and resistance to maize stripe virus (Dintinger et al. 2005; Jones et al. 2018). One of the resistance locus clusters (chr. 8) includes only QTL for two highly related fijiviruses in the family *Reoviridae*, suggesting this locus might also be unique.

Table 12.2 Overlapping virus resistance loci in maize

Chr[a]	Bin	Midpoint (Mb)[b]	Range (Mb)	Virus[c]	Family[d]	References
1	1.03	52.1 ± 3.2	41.1–67.7	MMV	*Rhabdoviridae*	Zambrano et al. (2014)
				MRCV	*Reoviridae*	DiRenzo et al. (2004)
2	2.02	60.6 ± 5.6	8.0–199.9	MRCV	*Reoviridae*	Martin et al. (2010)
				RBSDV	*Reoviridae*	Luan et al. (2012)
				MSpV	*Phenuiviridae*	Dinterger et al. (2005)
				MCMV	*Tombusviridae*	Jones et al. (2018)
	2.08	222.3 ± 5.1	211.2–231.5	MFSV	*Rhabdoviridae*	Zambrano et al. (2014)
				MMV	*Rhabdoviridae*	Zambrano et al. (2014)
				MSpV	*Phenuiviridae*	Dinterger et al. (2005)
3	3.05	122.5 ± 22.7	56.8–161.2	WSMV	*Potyviridae*	McMullen et al. (1994)
				SCMV	*Potyviridae*	Ding et al. (2012)
				MDMV	*Potyviridae*	Zambrano et al. (2014)
				MMV	*Rhabdoviridae*	Ming et al. (1998)
				MSpV	*Phenuiviridae*	Dintinger et al. (2005)
				WoMV	*Fimoviridae*	Lubberstedt et al. (2006)
				MSV	*Geminiviridae*	Welz et al. (1998)
				MCDV	*Secoviridae*	Jones et al. (2004)
4	4.08	212 ± 28.8	187.5–246.9	MRCV	*Reoviridae*	Bonamico et al. (2012)
				MSV	*Geminiviridae*	Welz et al. (1998)
				MCDV	*Secoviridae*	Jones et al. (2004)
6	6.01	27.3 ± 10.4	8.3–71.2	WSMV	*Potyviridae*	McMullen et al. (1994)
				MDMV	*Potyviridae*	Zambrano et al. (2014)
				SCMV	*Potyviridae*	Liu et al. (2017)
				MCDV	*Secoviridae*	Zambrano et al. (2014)
				WMoV	*Fimoviridae*	Lubberstedt et al. (2006)
				MFSV	*Rhabdoviridae*	Zambrano et al. (2014)
				MMV	*Rhabdoviridae*	Zambrano et al. (2014)
	6.05	153.5 ± 4.0	148–157	MSV	*Geminiviridae*	Pernet et al. (1999)
				MCMV	*Tombusviridae*	Jones et al. (2018)
8	8.07	173.1	168.6–173.1	RBSDV	*Reoviridae*	Luan et al. (2012)
				MRCV	*Reoviridae*	Bonamico et al. (2012)
10	10.05	130.4 ± 8.6	86.4–137.5	MRFV	*Tymoviridae*	Zambrano et al. (2014)
				MNeSV	*Tombusviridae*	Zambrano (2013)
				MCMV	*Tombusviridae*	Jones et al. (2018)
				BYDV	*Potyviridae*	Horn et al. (2015)
				WSMV	*Potyviridae*	McMullen et al. (1994)
				MDMV	*Potyviridae*	Zambrano et al. (2014)
				SCMV	*Potyviridae*	Zhang et al. (2013)
				MCDV	*Secoviridae*	Jones et al. (2004)
				MSV	*Geminiviridae*	Pernet et al. (1999)
				MSpV	*Phenuiviridae*	Dintinger et al. (2005)

[a]*Chr* chromosome
[b]Physical position in the B73 v3 genome
[c]*MMV* maize mosaic virus; *MRCV* Mal de Rio Cuarto virus; *RBSDV* rice black-streaked dwarf virus; *MSpV* maize stripe virus; *MCMV* maize chlorotic mottle virus; *MFSV* maize fine streak virus; *WSMV* wheat streak mosaic virus; *SCMV* sugarcane mosaic virus; *MDMV* maize dwarf mosaic virus; *WMoV* wheat mosaic virus; *MSV* maize streak virus; *MCDV* maize chlorotic dwarf virus; *MRFV* maize rayado fino virus; *MNeSV* maize necrotic streak virus
[d]*Family* the virus family

The major clusters of virus resistance loci include the regions of the potyvirus resistance clusters on chr. 3, 6, and 10 (Redinbaugh and Zambrano Mendoza 2014). These regions also carry loci for resistance to several fungal pathogens (Wisser et al. 2006) (Table 12.2). In addition to potyviruses, these clusters encode resistance to 5, 4, and 7 other viruses, respectively. Two clusters of virus resistance genes are present on chr. 2. One, in bin 2.02, includes QTL for resistance to three viruses in two different families (Di Renzo et al. 2004; Jones et al. 2018; Luan et al. 2012; Martin et al. 2010; Zambrano et al. 2014). In the other, overlapping QTL provides resistance for the rhabdoviruses MMV and MFSV and the tenuivirus MSpV (Dintinger et al. 2005; Zambrano et al. 2014). Similarly, clusters on chr. 4 and 6 (bin 6.05) confer resistance to two or three viruses in different virus families. The virus species within each cluster have little or no sequence identity, have different tissue specificities, and employ different replication and translation strategies.

Germplasm carrying strong resistance to one group of viruses in these clusters is not necessarily resistant to other types of viruses. For example, inbred line Pa405 is strongly resistant to potyviruses and WMoV; however, it is highly susceptible to a number of other viruses for which resistance loci are present on chr. 3, 6, and 10 including MCDV, MMV, MFSV, MNeSV, and MRFV. Because most resistant inbred lines used in these mapping studies carry resistance to a limited range of virus families, it seems likely that single loci are not responsible for providing resistance to all viruses within a cluster.

The inbred line Oh1VI was developed from an open-pollinated Virgin Island population as highly resistant to MCDV (Louie et al. 2002) and was subsequently found to be highly resistant to MDMV, SCMV, and WSMV (Jones et al. 2007). Further study indicated the line is highly resistant to MRFV, MMV, MFSV, and MNeSV and somewhat tolerant of MCMV (Zambrano et al. 2013; Mahuku et al. 2015; Jones et al. 2018). The virus resistance present in Oh1V1 was mapped to the same clusters previously identified in comparisons of mapping studies using diverse maize germplasm (Fig. 12.1). Further studies to fine-map virus resistance in these clusters in Oh1VI are ongoing. The results of these studies could provide opportunities to examine roles for a "birth and death" model, in which individual genes in multigene families are created by gene duplication and may later become inactivated or deleted from the genome, for virus resistance genes (Nei and Rooney 2005) or to identify unique mechanisms for virus resistance (Gomez et al. 2009).

12.4 Virus Resistance Genes in Maize

12.4.1 *Scmv1*

Following the identification of major resistance loci, *Scmv1* and *Scmv2*, tremendous efforts have been made to fine-map the two genes with mapping populations derived from the European

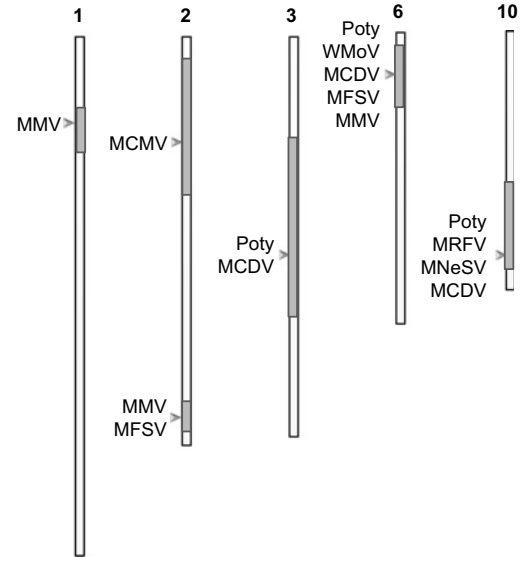

Fig. 12.1 Virus resistance in maize inbred line Oh1VI. The physical positions of markers associated with Oh1VI virus resistance QTL were determined by comparison to the B73 genome of research published. The positions of six gene clusters on five maize chromosomes are indicated by grey bars. The triangles indicate the mean position for identified QTL peaks. Potyvirus includes maize dwarf mosaic virus, sugarcane mosaic virus, and wheat streak mosaic virus

cross FAP1360A × F7 (Xu et al. 1999; Dussle et al. 2003; Yuan et al. 2004; Ingvardsen et al. 2010), Chinese maize inbred lines (Lü et al. 2008; Zhang et al. 2003), and tropical germplasm (Wu et al. 2007). Linkage mapping with three segregating populations finally assigned *Scmv1* to a 59.21 kb region of chr 6 containing three predicted genes (Tao et al. 2013). Candidate gene-based association mapping revealed that *ZmTrxH*, encoding an atypical h-type thioredoxin, was most likely to be the candidate for *Scmv1* (Tao et al. 2013; Leng et al. 2015). Inbred lines lacking the resistant allele of *ZmTrxH* were highly susceptible to SCMV. *ZmTrxH* was validated as *Scmv1* through a transgenic complementation assay, and *ZmTrxH* transcript abundance was demonstrated to be closely associated with resistance to SCMV. Intriguingly, *ZmTrxH* alleles from both resistant and susceptible lines shared identical coding/proximal promoter regions, but varied in their upstream regulatory regions. In contrast to more than 30 other thioredoxins encoded by the maize genome, *ZmTrxH* has an atypical WNQPS structure within the thioredoxin active-site motif, in which the two canonical cysteines found in the other thioredoxins are replaced by asparagine (N) and serine (S) in both the resistant and susceptible alleles (Liu et al. 2017). This change renders *ZmTrxH* unable to reduce disulfide bridges, the typical activity of thioredoxins; however, the *ZmTrxH* protein has a strong molecular chaperone-like activity. Thioredoxins have previously been implicated in virus infection and resistance, with the silencing of a maize m-type thioredoxin enhancing systemic infection of SCMV (Shi et al. 2011). In addition, overexpression of a *Nicotiana benthamaniana* h-type thioredoxin conferred resistance to tobacco mosaic virus and cucumber mosaic virus, two (+)-strand RNA viruses from different families (Sun et al. 2010). *ZmTrxH* is dispersed in the cytoplasm and suppresses viral accumulation without eliciting the SA- or JA-mediated pathogen defense signaling pathway associated with R-gene-mediated resistance (Liu et al. 2017). These results shed new insight into plant viral defense mechanisms and define a process which is obviously different from that conferred by NB-LRR-type R genes.

12.4.2 *Scmv2*

Using a large isogenic mapping population, *Scmv2* was mapped to an interval of 1.34 Mb on chr 3, covering four predicted genes possibly involved in virus movement (Ingvardsen et al. 2010). Later, *Scmv2* was fine-mapped to an interval of 196.5 kb with two predicted genes, encoding an auxin-binding protein (ABP1) and a Rho GTPase-activating protein, as candidate genes for *Scmv2* (Ding et al. 2012). Candidate gene-based association mapping revealed a significantly associated marker 207FG003 in the *Scmv2* region (Leng et al. 2015). Combined genome-wide association study (GWAS) and linkage analyses revealed four genes at *Scmv2*, and one of them, encoding *ABP1*, was the most likely candidate for *Scmv2* (Li et al. 2016). The native *ABP1* gene, including 1.7 kb of the promoter region and 1 kb downstream of the coding region, was isolated from the resistant line FAP1360A and used for functional complementation assay (Leng et al. 2017). Susceptible genotypes at the *Scmv2* locus were complemented by transgenic full-length *ABP1* to confer resistance, while downregulation of *ABP1* by RNAi resulted in susceptible plants. Sequence variation in the *ABP1* promoter region resulted in higher expression that was associated with SCMV resistance. The ABP1 protein has no effect on SCMV replication, but it most likely confines systemic viral infection by directly interacting with Rubisco small subunit (RbCS) (Leng et al. 2017). Thus, the well-characterized gene *ABP1* confers resistance to a potyvirus in plants.

Previous studies have implicated interactions between the chloroplast and viral proteins in the development of disease. In particular, interactions between potyvirus coat proteins and chloroplast components have been identified (reviewed in Zhao et al. 2016). The RbCS protein has been implicated in resistance to tomato mosaic virus conferred by the *Tm-2*^{2} gene

product, which interacts with the viral movement protein to prevent systemic virus movement (Zhao et al. 2013).

12.4.3 Other Potential Virus Resistance Genes

Recessive virus resistance genes in plants have previously been associated with mutations in translation factors (Robaglia and Caranta 2006), and recessive alleles of eIF4e confer virus resistance in several hosts (Diaz-Pendon et al. 2004; Gomez et al. 2009). In this resistance mechanism, the protein produced from the recessive allele fails to interact with the virus and recruit the viral RNA to cap-binding complex. In addition, eIF4e has been shown to move from cell to cell, with some alleles of eIF4e preventing cell-to-cell movement of the potyvirus, pea seed-borne mosaic virus (Gao et al. 2004). Within the larger genome region of chr. 3 that carries QTL for resistance to MCDV, MMV, and MRFV in the multi-virus-resistant inbred line Oh1VI, the B73 genome encodes two genes for the translation factor eIF4e (Zambrano et al. 2014). However, these eIF4e genes are not within the *Scmv2* regions identified in fine-mapping studies (Ding et al. 2012; Ingvardsen et al. 2010; Leng et al. 2017). The recessive character of the MCMV tolerance QTL that mapped to chr 6 in maize inbreds KS23-5 and KS23-6 could be associated with a translation factor. Genes encoding elongation factor 1 alpha (eEF1A) are present on chr 6. In several plant virus systems, eEF1A interact with viral replicases and are thought to recruit viral RNAs to the replication complex (Sanfacon 2015). Further research is needed to determine whether translation factors play any role in virus resistance in maize.

Although other virus resistance genes remain to be identified from maize, pathogen-derived resistance in transgenic maize expressing viral RNAs derived from MCDV, SCMV, and MSV has resistance to these viruses (Liu et al. 2009a; McMullen et al. 1996; Shepherd et al. 2007; Shepherd et al. 2014; Zhang et al. 2013). In addition, maize expressing an *E. coli* ribonuclease specific for double-stranded RNA had increased resistance to RBSDV (Cao et al. 2013).

12.5 Genomic and Transcriptomic Responses to Viruses in Maize

Information on the responses of susceptible and resistant maize to virus inoculation is accumulating. In dicots, inoculation with viruses has been shown to increase cellular stress and defense gene expression, alter expression of genes regulating development and hormone responses, and increase expression of genes involved in RNAi (reviewed in Whitham et al. 2006). For maize, the transcriptomic and proteomic responses of resistant and susceptible plants to infection with SCMV, MDMV, and RBSDV have been studied. Despite the differences in these viruses—the potyvirus (SCMV and MDMV) genomes are monopartite, single-stranded, positive sense RNA, and the RBSDV genome is multipartite, enveloped, double-stranded RNA—there are common themes in the responses of susceptible and resistant maize. Similar to dicots, increased levels of defense genes were noted in maize inoculated with potyviruses or RBSDV up to 9 and at 50-day post-inoculation, respectively (Cassone et al. 2014; Jia et al. 2012; Li et al. 2011; Shi et al. 2006; Uzarowska et al. 2009; Wu et al. 2013a; Zhou et al. 2016). However, differences in both specific transcripts/proteins that accumulated and the timing of their accumulation were noted between resistant and susceptible maize inbreds, with the responses generally being of greater magnitude and/or faster in resistant plants. Other virus-related changes included expression of genes associated with carbohydrate and energy metabolism, protein degradation, signal transduction, hormone synthesis and response, and cell wall development.

Transcripts of genes with functions in RNA interference (RNAi), the pathways used by many organisms to regulate gene expression and virus infections, accumulated in both RBSDV- and

SCMV-inoculated maize. In experiments to characterize siRNA associated with RBSDV infection, inoculation of susceptible maize resulted in accumulation of gene-specific transcripts for dicer (*Dcl1*, *2, 3a*), argonaute (*Ago1a, 1b2, 18a*) and RNA-dependent RNA polymerase (*Rdr6*) (Li et al. 2017). Experiments to characterize siRNA in susceptible maize after SCMV inoculation revealed upregulation of *Dcl2* and *Ago2*, but downregulation of *Dcl4* (Xia et al. 2014). *Dcl2* accumulated in susceptible maize inoculated with SCMV, MCMV or both viruses (MLN), with significantly higher levels of transcript in MLN-inoculated plants, but expression of other *Dcl* genes was either not affected or reduced by virus inoculation (Xia et al. 2016). In this system, *Ago2A* and *Ago18a* accumulated in virus inoculated plants, with patterns similar to *Dcl2*. In contrast, the highest levels of *Ago1a*, *1b*, and *1c* transcripts were found in SCMV-inoculated plants (Li et al. 2017). In Arabidopsis, *Ago2* and *Dcl2* are required to control viral infections caused by adapted viruses (Zhang et al. 2012). Although the roles of specific *Dcl* and *Ago* genes have not been defined in maize, *Dcl2* appears to be required for efficient intercellular movement of the virus-induced gene silencing (VIGS) signal in *N. benthamiana* (Qin et al. 2017). While at least some members of the *Dcl*, *Ago* and *Rdr* families co-localize with the observed clusters of virus resistance loci in maize, the relationship between these genes and resistance is not known.

Increased expression of transcripts for genes important for photosynthesis has been noted in some systems. For example, the large subunit of RuBisCO accumulated in resistant maize inoculated with SCMV (Wu et al. 2013a; Wu et al. 2015). Chloroplast localized ferredoxin V and thioredoxins was also upregulated (Cao et al. 2012; Cheng et al. 2008; Wu et al. 2015). Taken together with interactions of SCMV viral proteins with the RuBisCo small subunit and ferredoxin (Cheng et al. 2008; Leng et al. 2017), the results suggest an intimate relationship between photosynthetic activity and potyvirus infection.

It is perhaps not surprising that similar regulation of only a very limited number of specific genes or proteins was identified in the experiments outlined above, because of differences in the viruses, germplasm (including the presence or absence of resistance), time after inoculation and even the tissues analyzed. Among the common threads were increased accumulation of β-glucanase transcripts in plants inoculated with either SCMV or MDMV, with higher levels in resistant lines (Cassone et al. 2014; Shi et al. 2006; Uzarowska et al. 2009). Increased accumulation of transcripts with similarity to brassinosteroid-insensitive receptor kinase, a gene with roles in innate immunity and plant growth, occurred in virus-resistant plants inoculated with MDMV and RBSDV (Cassone et al. 2014; Huot et al. 2014; Jia et al. 2012). Remorin genes were upregulated in resistant and susceptible MDMV-inoculated plants Cassone et al. 2014, and susceptible SCMV-inoculated plants (Wu et al. 2013b). Remorin proteins are located within punctate membrane microdomains and have been implicated in virus spread in plants (Konrad and Ott 2015; Raffaele et al. 2009). While none of these genes has been associated with a specific virus resistance QTL, the similar regulation of genes provides a basis for development of studies to examine the roles of specific genes and pathways in virus resistance and susceptibility in maize.

Changes in the expression of specific genes have been associated with virus resistance in maize. As noted above, 100-fold higher expression of the *Scmv1/ZmTrxH* gene was associated with the resistance response in line FAP1360A relative to susceptible controls (Liu et al. 2017). Interestingly, an m-type thioredoxin mapping to chr. 5 was also upregulated in maize inoculated with SCMV, and silencing of its expression inhibited SCMV accumulation in maize and tobacco vein-banding virus in tobacco (Shi et al. 2011). Cao et al. (2012) showed that a Rho-related GTPase induced during SCMV infection of susceptible plants is required for virus infection. Rop genes have been shown to regulate pathogen resistance including virus resistance (Sacco et al. 2007; Zhang et al. 2014) and have been implicated in abscisic acid responses, development and stress responses

(Craddock et al. 2012). Elongin C, a transcription factor that increases transcription elongation by RNA polymerase II, interacts with the potyviral genomic protein and is expressed at higher levels at 4–6 days post-inoculation (Zhu et al. 2014). In addition, the SCMV HC-Pro interacts with the transit peptide of the chloroplastic ferredoxin V, and expression of this gene is downregulated during SCMV infection (Cheng et al. 2008). The exact roles of these proteins in enhancing or suppressing virus infection and their association with QTL for virus resistance remain to be determined.

12.6 Conclusions

Improvements in phenotyping plants for virus resistance, genotyping maize populations, and functional analysis of candidate genes are likely to accelerate increased understanding of genes, proteins, and mechanisms associated with virus resistance in maize. The recent identification of the genes underlying *Scmv1* and *Scmv2* provides the basis for understanding whether potyvirus resistance, and resistance to other viruses, is pleiotropic in maize. Characterization of these genes will also aid in our understanding of the mechanisms some isolates use to break resistance that is critical to understanding the durability of alleles deployed to control the disease. Our ability to edit plant genomes will facilitate validation of the importance of candidate genes in virus resistance and should aid in the development of highly productive, disease-resistant maize crops.

References

Adams IP, Miano DW, Kinyua ZM, Wangai A, Kimani E, Phiri N, Reeder R, Harju V, Glover R, Hany U, Souza-Richards R, Nath PD, Nixon T, Fox A, Barnes A, Smith J, Skelton A, Thwaites R, Mumford R, Boonham N (2013) Use of next-generation sequencing for the identification and characterization of Maize chlorotic mottle virus and sugarcane mosaic virus causing maize lethal necrosis in Kenya. Plant Pathol 62:741–749. https://doi.org/10.1111/j.1365-3059.2012.02690.x

Albrechtsen SE (2006) Seed transmission of viruses. CABI Publishing-C A B INT. https://doi.org/10.1079/9780851990163.0013

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol Evol 19:535–544

Bernreiter A, Garcia Teijeiro R, Jarrin D, Garrido P, Ramos L (2017) First report of Maize yellow mosaic virus infecting maize in Ecuador. New Dis Rep 11. http://dx.doi.org/10.5197/j.2044-0588.2017.036.011

Bonamico NC, Balzarini MG, Arroyo AT, Ibanez MA, Diaz DG, Salerno JC, Renzo MA (2010) Association between microsatellites and resistance to Mal de Rio Cuarto in maize by discriminant analysis. Phyton-Int J Exp Bot 79:31–38

Bonamico NC, Di Renzo MA, Ibanez MA, Borghi ML, Diaz DG, Salerno JC, Balzarini MG(2012) QTL analysis of resistance to Mal de Rio Cuarto disease in maize using recombinant inbred lines. J Agric Sci 150:619–629

Boulton MI, Buchholz WG, Marks MS, Markham PG, Davies JW (1989) Specificity of *Agrobacterium*-mediated delivery of maize streak virus-DNA to members of the Gramineae. Plant Mol Biol 12:31–40

Brewbaker JL (1979) Diseases of maize in the wet lowland tropics and the collapse of the classic Maya civilization. Econ Bot 33:101–118

Brewbaker JL (1981) Resistance to maize mosaic virus. In: Gordon DT, Knoke JK, Scott GE (eds) Virus and virus-like diseases of Maize in the United States. So Coop Ser Bull, Wooster, OH, pp 145–151

Cabanas D, Watanabe S, Higashi CHV, Bressan A (2013) Dissecting the mode of *Maize chlorotic mottle virus* transmission (*Tombusviridae: Machlomovirus*) by *Frankliniella williamsi* (Thysanoptera: Thripidae). J Econ Entomol 106:16–24. https://doi.org/10.1603/ec12056

Cao XL, Lu YG, Di DP, Zhang ZY, Liu H, Tian LZ, Zhang AH, Zhang YJ, Shi LD, Guo BH, Xu J, Duan XF, Wang XB, Han CG, Miao HQ, Yu JL, Li DW (2013) Enhanced virus resistance in transgenic maize expressing a dsRNA-specific endoribonuclease gene from *E. coli*. Plos One 8:e60829. https://doi.org/10.1371/journal.pone.0060829

Cao YY, Shi Y, Li YQ, Cheng YQ, Zhou T, Fan ZF (2012) Possible involvement of maize Rop1 in the defense responses of plants to viral infection. Mol Plant Pathol 13:732–743. https://doi.org/10.1111/j.1364-3703.2011.00782.x

Cassone BJ, Chen ZB, Chiera J, Stewart LR, Redinbaugh MG (2014) Responses of highly resistant and susceptible maize to vascular puncture inoculation with Maize dwarf mosaic virus. Physiol Mol Plant Pathol 86:19–27. https://doi.org/10.1016/j.pmpp.2014.01.002

Chen S, Cao L, Huang Q, Qian Y, Zhou X (2016) The complete genome sequence of a novel maize-associated totivirus. Adv Virol 161:487–490. https://doi.org/10.1007/s00705-015-2657-y

Cheng YQ, Liu ZM, Xu J, Zhou T, Wang M, Chen YT, Li HF, Fan ZF (2008) HC-Pro protein of sugar cane mosaic virus interacts specifically with maize ferredoxin-5 *in vitro* and *in planta*. J Gen Virol 89:2046–2054. https://doi.org/10.1099/vir.0.2008/001271-0

Craddock C, Lavagi I, Yang Z (2012) New Insights into Rho signaling from plant ROP/Rac GTPases. Trends Cell Biol 22:492–501. https://doi.org/10.1016/j.tcb.2012.05.002

Deng TC, Chou CM, Chen CT, Tsai CH, Lin FC (2014) First report of Maize chlorotic mottle virus on sweet corn in Taiwan. Plant Dis 98:1748. https://doi.org/10.1094/PDIS-06-14-0568-PDN

Di Renzo MA, Bonamico NC, Diaz DG, Ibanez MA, Faricelli ME, Balzarini MG, Salerno JC (2004) Microsatellite markers linked to QTL for resistance to Mal de Rio Cuarto disease in *Zea mays* L. J Agric Sci 142:289–295

Diaz-Pendon JA, Truniger V, Nieto C, Garcia-Mas J, Bendahmane A, Aranda MA (2004) Advances in understanding recessive resistance to plant viruses. Mol Plant Pathol 5:223–233

Ding JQ, Li HM, Wang YX, Zhao RB, Zhang XC, Chen JF, Xia ZL, Wu JY (2012) Fine mapping of *Rscmv2*, a major gene for resistance to sugarcane mosaic virus in maize. Mol Breed 30:1593–1600. https://doi.org/10.1007/s11032-012-9741-8

Dintinger J, Verger D, Caiveau S, Risterucci AM, Gilles J, Chiroleu F, Courtois B, Reynaud B, Hamon P (2005) Genetic mapping of maize stripe disease resistance from the Mascarene source. Theor Appl Genet 111:347–359

Dussle CM, Quint M, Melchinger AE, Xu ML, Lubberstedt T (2003) Saturation of two chromosome regions conferring resistance to SCMV with SSR and AFLP markers by targeted BSA. Theor Appl Genet 106:485–493

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. e1937910.1371/journal.pone.0019379

Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Ganal MW, Durstewitz G, Polley A, Berard A, Buckler ES, Charcosset A, Clarke JD, Graner EM, Hansen M, Joets J, Le Paslier MC, McMullen MD, Montalent P, Rose M, Schon CC, Sun Q, Walter H, Martin OC, Falque M (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6:e28334. e2833410.1371/journal.pone.0028334

Gao ZH, Johansen E, Eyers S, Thomas CL, Noel Ellis TH, Maule AJ (2004) The potyvirus recessive resistance gene, sbm1, identifies a novel role for translation initiation factor eIF4E in cell-to- cell trafficking. Plant J 40:376–385

Gomez P, Rodriguez-Hernandez AM, Moury B, Aranda MA (2009) Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. Eur J Plant Pathol 125:1–22. https://doi.org/10.1007/s10658-009-9468-5

Herold F, Bergold GH, Weibel J (1960) Isolation and electron microscopic demonstration of a virus infecting corn (Zea mays L.). Virology 12:335–347

Horn F, Habekuss A, Stich B (2015) Linkage mapping of barley yellow dwarf virus resistance in connectedpopulations of maize. BMC Plant Biol 15:29

Hull R (2002) Chapter 1—Introduction. In: Matthews' plant virology, 4th edn. Academic Press, London, pp 1–12. https://doi.org/10.1016/B978-012361160-4/50052-9

Huot B, Yao J, Montgomery BL, He SY (2014) Growth-defense tradeoffs in plants: a balancing act to optimize fitness. Mol Plant 7:1267–1287. https://doi.org/10.1093/mp/ssu049

Ingvardsen CR, Xing YZ, Frei UK, Lubberstedt T (2010) Genetic and physical fine mapping of *Scmv2*, a potyvirus resistance gene in maize. Theor Appl Genet 120:1621–1634. https://doi.org/10.1007/s00122-010-1281-1

Jensen SG, Lane LC, Seifers DL (1996) A new disease of maize and wheat in the High Plains. Plant Dis 80:1387–1390

Jia MA, Li YQ, Lei L, Di DP, Miao HQ, Fan ZF (2012) Alteration of gene expression profile in maize infected with a double-stranded RNA fijivirus associated with symptom development. Mol Plant Pathol 13:251–262. https://doi.org/10.1111/j.1364-3703.2011.00743.x

Johansen E, Edwards MC, Hampton RO (1994) Seed transmission of viruses—current perspectives. Annu Rev Phytopathol 32:363–386

Jones M, Boyd E, Redinbaugh M (2011) Responses of maize (*Zea mays* L.) near isogenic lines carrying *Wsm1*, *Wsm2*, and *Wsm3* to three viruses in the *Potyviridae*. Theor Appl Genet 123:729–740. https://doi.org/10.1007/s00122-011-1622-8

Jones MW, Penning BW, Jamann TM, Glaubitz JC, Romay MC, Buckler ES, Redinbaugh MG (2018) Diverse chromosomal locations of quantitative trait loci for tolerance to maize chlorotic mottle virus in five maize populations. Phytopathology (in press)

Jones MW, Redinbaugh MG, Louie R (2007) The *Mdm1* locus and maize resistance to *Maize dwarf mosaic virus*. Plant Dis 91:185–190

Jones MW, Redinbaugh MG, Anderson RJ, Louie R (2004) Identification of quantitative trait locicontrolling resistance to maize chlorotic dwarf virus. Theor Appl Genet 110:48–57

Kang BC, Yeam I, Jahn MM (2005) Genetics of plant virus resistance. Annu Rev Phytopathol 43:581–621

Konrad SSA, Ott T (2015) Molecular principles of membrane microdomain targeting in plants. Trends Plant Sci 20:351–361. https://doi.org/10.1016/j.tplants.2015.03.016

Kuntze L, Fuchs E, Gruntzig M, Schulz B, Klein D, Melchinger AE (1997) Resistance of early-maturing European maize germplasm to *sugarcane mosaic virus* (SCMV) and *maize dwarf mosaic virus* (MDMV). Plant Breed 116:499–501

Lapierre H, Signoret P (2004) Virus and virus diseases of *Poaceae (Gramineae)*. Mieux comprendre. Institut National de la Recherche Agronomique, Paris. ISSN 1144-7605

Leng P, Ji Q, Tao Y, Ibrahim R, Pan G, Xu M, Lübberstedt T (2015) Characterization of sugarcane mosaic virus *Scmv1* and *Scmv2* resistance regions by regional association analysis in maize. PLoS ONE 10:e0140617. https://doi.org/10.1371/journal.pone.0140617

Leng P, Ji Q, Asp T, Frei UK, Ingvardsen CR, Xing Y, Studer B, Redinbaugh M, Jones M, Gajjar P, Liu S, Li F, Pan G, Xu M, Lübberstedt T (2017) Auxin binding protein 1 reinforces resistance to sugarcane mosaic virus in maize. Mol Plant 10:1357–1360. https://doi.org/10.1016/j.molp.2017.07.013

Li K, Xu C, Zhang J (2011) Proteome profile of maize (*Zea mays* L.) leaf tissue at the flowering stage after long-term adjustment to *rice black-streaked dwarf virus* infection. Gene 485:106–113. https://doi.org/10.1016/j.gene.2011.06.016

Li MJ, Li YQ, Xia ZH, Di DP, Zhang AH, Miao HQ, Zhou T, Fan ZF (2017) Characterization of small interfering RNAs derived from *Rice black streaked dwarf virus* in infected maize plants by deep sequencing. Virus Res 228:66–74. https://doi.org/10.1016/j.virusres.2016.11.001

Li Z, Chen J, Han L, Wen J, Chen G, Li H, Wang Y, Zhao R, Zhang X, Xia Z, Yan J, Wu J, Ding J (2016) Association mapping resolving the major loci *Scmv2* conferring resistance to sugarcane mosaic virus in maize. Eur J Plant Pathol 145:385–391. https://doi.org/10.1007/s10658-015-0852-z

Liu Q, Liu H, Gong Y, Tao Y, Jiang L, Zuo W, Yang Q, Ye J, Lai J, Wu J, Lübberstedt T, Xu M (2017) An atypical thioredoxin imparts early resistance to *sugarcane mosaic virus* in maize. Mol Plant 10:483–497. https://doi.org/10.1016/j.molp.2017.02.002

Liu XH, Tan ZB, Li WC, Zhang HM, He DW (2009a) Cloning and transformation of SCMV CP gene and regeneration of transgenic maize plants showing resistance to SCMV strain MDB. Afr J Biotechnol 8:3747–3753

Liu XH, Tan ZB, Rong TZ (2009b) Molecular mapping of a major QTL conferring resistance to SCMV based on immortal RIL population in maize. Euphytica 167:229–235. https://doi.org/10.1007/s10681-008-9874-3

Louie R, Findley WR, Knoke JK, McMullen MD (1991) Genetic-basis of resistance in maize to 5 *Maize dwarf mosaic virus* strains. Crop Sci 31:14–18

Louie R (1995) Vascular puncture of maize kernels for the mechanical transmission of maize white line mosaic virus and other viruses of maize. Phytopathology 85:139–143

Louie R, Jones MW, Anderson RJ, Redinbaugh MG (2002) Registration of maize germplasm Oh1VI. Crop Sci 42:991

Lü X, Li X, Xie C, Hao Z, Ji H, Shi L, Zhang S (2008) Comparative QTL mapping of resistance to sugarcane mosaic virus in maize based on bioinformatics. Front Agric China 2:365–371. https://doi.org/10.1007/s11703-008-0081-8

Luan JW, Wang F, Li YJ, Zhang B, Zhang JR (2012) Mapping quantitative trait loci conferring resistance to rice black-streaked virus in maize (Zea mays L.). Theor Appl Genet 125:781–791. https://doi.org/10.1007/s00122-012-1871-1

Lubberstedt T, Ingvardsen C, Melchinger AE, Xing Y, Salomon R, Redinbaugh MG (2006) Two chromosome segments confer multiple potyvirus resistance in maize. Plant Breed 125:352–356

Mahuku G, Lockhart BE, Wanjala B, Jones MW, Kimunye JN, Stewart LR, Cassone BJ, Sevgan S, Nyasani JO, Kusia E, Kumar PL, Niblett CL, Kiggundu A, Asea G, Pappu HR, Wangai A, Prasanna BM, Redinbaugh MG (2015) Maize lethal necrosis (MLN), an emerging threat to maize-based food security in Sub-Saharan Africa. Phytopathology 105:956–965. https://doi.org/10.1094/PHYTO-12-14-0367-FI

Martin DP, Shepherd DN (2009) The epidemiology, economic impact and control of maize streak disease. Food Secur 1:305–315. https://doi.org/10.1007/s12571-009-0023-1

Martin T, Franchino JA, Kreff ED, Procopiuk AM, Tomas A, Luck S, Shu GG (2010) Major QTLs conferring resistance of corn to Fijivirus. U.S. Patent US 2010/0325750 Al

McMullen MD, Louie R (1989) The linkage of molecular markers to a gene controlling the symptom response in maize to *Maize dwarf mosaic virus*. Mol Plant-Microbe Interact 2:309–314

McMullen MD, Jones MW, Simcox KD, Louie R (1994) 3 genetic-loci control resistance to *wheat streak mosaic virus* in the maize inbred Pa405. Mol Plant-Microbe Interact 7:708–712

McMullen MD, Roth BA, Townsend R (1996) Maize chlorotic dwarf virus and resistance thereto. Patent, U.S.A

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li HH, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Rosas MO, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740. https://doi.org/10.1126/science.1174320

Melchinger AE, Kuntze L, Gumber RK, Lübberstedt T, Fuchs E (1998) Genetic basis of resistance to *sugarcane mosaic virus* in European maize germplasm. Theor Appl Genet 96:1151–1161

Nair SK, Babu R, Magorokosho C, Mahuku G, Semagn K, Beyene Y, Das B, Makumbi D, Kumar PL, Olsen M, Boddupalli PM (2015) Fine mapping of *Msv1*, a major QTL for resistance to maize streak virus leads to development of production markers for breeding pipelines. Theor Appl Genet 128:1839–1854. https://doi.org/10.1007/s00122-015-2551-8

Nault LR, Styer WE, Coffey ME, Gordon DT, Negi LS, Niblett CL (1978) Transmission of maize chlorotic mottle virus by Chrysomelid beetles. Phytopathology 68:1071–1074

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152. https://doi.org/10.1146/annurev.genet.39.073003.112240

Niblett CL, Claflin LE (1978) Corn lethal necrosis—new virus disease of corn in Kansas. Plant Dis Report 62:15–19

Oerke EC, Dehne HW (2004) Safeguarding production—losses in major crops and the role of crop protection. Crop Prot 23:275–285. https://doi.org/10.1016/j.cropro.2003.10.001

Palanga E, Longué RDS, Koala M, Néya JB, Traoré O, Martin DP, Peterschmitt M, Filloux D, Roumagnac P (2017) First report of maize yellow mosaic virus infecting maize in Burkina Faso. New Dis Report 35:26. http://dx.doi.org/10.1094/PDIS-03-17-0315-PDN

Pernet A, Hoisington D, Franco J, Isnard M, Jewell D, Jiang C, Marchand JL, Reynaud B, Glaszmann JC, de Leon DG (1999) Genetic mapping of maize streak virus resistance from the Mascarene source. I.Resistance in line D211 and stability against different virus clones. Theor Appl Genet 99:524–539

Pratt RC, Gordon SG (2006) Breeding for resistance to maize foliar pathogens. Plant Breed Rev 26:119–173

Qin C, Li B, Fan Y, Zhang X, Yu Z, Ryabov E, Zhao M, Wang H, Shi N, Zhang P, Jackson S, Tör M, Cheng Q, Liu Y, Gallusci P, Hong Y (2017) Roles of Dicer-like proteins 2 and 4 in intra- and intercellular antiviral silencing. Plant Physiol 174:1067–1081

Quito-Avila DF, Alvarez RA, Mendoza AA (2016) Occurrence of maize lethal necrosis in Ecuador: a disease without boundaries? Eur J Plant Pathol 146:705–710. https://doi.org/10.1007/s10658-016-0943-5

Raffaele S, Bayer E, Lafarge D, Cluzet S, Retana SG, Boubekeur T, Leborgne-Castel N, Carde JP, Lherminier J, Noirot E, Satiat-Jeunemaitre B, Laroche-Traineau J, Moreau P, Ott T, Maule AJ, Reymond P, Simon-Plas F, Farmer EE, Bessoule JJ, Mongrand S (2009) Remorin, a Solanaceae protein resident in membrane rafts and plasmodesmata, impairs *Potato virus X* movement. Plant Cell 21:1541–1555. https://doi.org/10.1105/tpc.108.064279

Ranum P, Peña-Rosas JP, Garcia-Casal MN (2014) Global maize production, utilization, and consumption. Anna New York Acad Sci 1312:105–112. https://doi.org/10.1111/nyas.12396

Redinbaugh MG, Pratt RC (2008) Virus Resistance. In: Hake S, Bennetzen JL (eds) Handbook of maize: its biology. Springer-Verlag, New York, pp 255–270

Redinbaugh MG, Zambrano Mendoza JL (2014) Control of virus diseases in maize. In: Lobenstein G, Katis N (eds) Control of plant virus disease: seed-propogated crops. Advances in Virus Research, vol 90. Academic Press, Waltham, MA USA, pp 391–429

Robaglia C, Caranta C (2006) Translation initiation factors: a weak link in plant RNA virus infection. Trends Plant Sci 11:40–45. https://doi.org/10.1016/j.tplants.2005.11.004

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES, Gardner CA (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14:R55. https://doi.org/10.1186/gb-2013-14-6-r55

Sacco MA, Mansoor S, Moffett P (2007) A RanGAP protein physically interacts with the NB-LRR protein Rx, and is required for Rx-mediated viral resistance. Plant J 52:82–93. https://doi.org/10.1111/j.1365-313X.2007.03213.x

Sanfacon H (2015) Plant translation factors and virus resistance. Viruses-Basel 7:3392–3419. https://doi.org/10.3390/v7072778

Shepherd DN, Mangwende T, Martin DP, Bezuidenhout M, Kloppers FJ, Carolissen CH, Monjane AL, Rybicki EP, Thomson JA (2007) Maize streak virus-resistant transgenic maize: a first for Africa. Plant Biotechnol J 5:759–767

Shepherd DN, Dugdale B, Martin DP, Varsani A, Lakay FM, Bezuidenhout ME, Monjane AL, Thomson JA, Dale J, Rybicki EP (2014) Inducible resistance to maize streak virus. Plos One 9:e105932. https://doi.org/10.1371/journal.pone.0105932

Shi C, Thummler F, Melchinger AE, Wenzel G, Lubberstedt T (2006) Comparison of transcript profiles between near-isogenic maize lines in association with SCMV resistance based on unigene-microarrays. Plant Sci 170:159–169

Shi Y, Qin YH, Cao YY, Sun H, Zhou T, Hong YG, Fan ZF (2011) Influence of an m-type thioredoxin in maize on potyviral infection. Eur J Plant Pathol 131:317–326. https://doi.org/10.1007/s10658-011-9810-6

Stenger D, Hein G, Tatineni S, French R (2016) Chapter 18: Eriophyid Mite Vectors of Plant Viruses. In: Brown JK (ed) Vector-mediated transmission of plant pathogens. General plant pathology. The American Phytopathological Society, pp 263–274. https://doi.org/10.1094/9780890545355.01810.1094/9780890545355.018

Stewart LR, Haque MA, Jones MW, Redinbaugh MG (2013) Response of maize (*Zea mays* L.) lines

carrying *Wsm1*, *Wsm2*, and *Wsm3* to the potyviruses *Johnsongrass mosaic virus* and *sorghum mosaic virus*. Mol Breed 31:289–297. https://doi.org/10.1007/s11032-012-9789-5

Stewart LR, Louie R, Redinbaugh MG (2016) Virus diseases. In: Munkvold GPaW, DG (ed) Compendium of corn diseases, 4th edn. APS Press, St. Paul, MN, pp 100–116

Stewart LR, Willie K, Wijeratne S, Redinbaugh MG, Massawe D, Niblett CL, Kiggundu A, Assimwe T (2017) Johnsongrass mosaic virus contributes to Maize lethal necrosis in East Africa. Plant disease (in press)

Sun LJ, Ren HY, Liu RX, Li BY, Wu TQ, Sun F, Liu HM, Wang XM, Dong HS (2010) An h-type thioredoxin functions in tobacco defense responses to two species of viruses and an abiotic oxidative stress. Mol Plant-Microbe Interact 23:1470–1485. https://doi.org/10.1094/mpmi-01-10-0029

Tao YF, Jiang L, Liu QQ, Zhang Y, Zhang R, Ingvardsen CR, Frei UK, Wang BB, Lai JS, Lubberstedt T, Xu ML (2013) Combined linkage and association mapping reveals candidates for Scmv1, a major locus involved in resistance to sugarcane mosaic virus (SCMV) in maize. BMC Plant Biol 13:162. https://doi.org/10.1186/1471-2229-13-162

Uzarowska A, Dionisio G, Sarholz B, Piepho HP, Xu ML, Ingvardsen CR, Wenze G, Lubberstedt T (2009) Validation of candidate genes putatively associated with resistance to SCMV and MDMV in maize (*Zea mays* L.) by expression profiling. BMC Plant Biol 9:15. https://doi.org/10.1186/1471-2229-9-15

Wang F, Zhou BG, Gao ZL, Xu DF (2016) A new species of the genus *Polerovirus* causing symptoms similar to *Maize yellow dwarf virus*-RMV of maize in China. Plant Dis 100:1508–1508. https://doi.org/10.1094/pdis-11-15-1259-pdn

Wangai AW, Redinbaugh MG, Kinyua ZM, Miano DW, Leley PK, Kasina M, Mahuku G, Scheets K, Jeffers D (2012) First Report of *Maize chlorotic mottle virus* and Maize Lethal Necrosis in Kenya. Plant Dis 96:1582–1583. https://doi.org/10.1094/pdis-06-12-0576-pdn

Welz HG, Schechert A, Pernet A, Pixley KV, Geiger HH (1998) A gene for resistance to the maize streak virus in the African CIMMYT maize inbred line CML202. Mol Breed 4:147–154

Whitham SA, Yang CL, Goodin MM (2006) Global impact: Elucidating plant responses to viral infection. Mol Plant-Microbe Interact 19:1207–1215

Wisser RJ, Balint-Kurti PJ, Nelson RJ (2006) The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96:120–129

Wu JY, Ding JQ, Du YX, Xu YB, Zhang XC (2007) Genetic analysis and molecular mapping of two dominant complementary genes determining resistance to sugarcane mosaic virus in maize. Euphytica 156:355–364

Wu L, Han Z, Wang S, Wang X, Sun A, Zu X, Chen Y (2013a) Comparative proteomic analysis of the plant-virus interaction in resistant and susceptible ecotypes of maize infected with *sugarcane mosaic virus*. J Proteomics 89:124–140. https://doi.org/10.1016/j.jprot.2013.06.005

Wu L, Wang S, Wu J, Han Z, Wang R, Wu L, Zhang H, Chen Y, Hu X (2015) Phosphoproteomic analysis of the resistant and susceptible genotypes of maize infected with *sugarcane mosaic virus*. Amino Acids 47:483–496. https://doi.org/10.1007/s00726-014-1880-2

Wu LJ, Han ZP, Wang SX, Wang XT, Sun AG, Zu XF, Chen YH (2013b) Comparative proteomic analysis of the plant-virus interaction in resistant and susceptible ecotypes of maize infected with sugarcane mosaic virus. J Proteomics 89:124–140. https://doi.org/10.1016/j.jprot.2013.06.005

Xia XC, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to *Sugarcane mosaic virus* in maize. Phytopathology 89:660–667

Xia Z, Zhao ZX, Chen L, Li MJ, Zhou T, Deng CL, Zhou Q, Fan ZF (2016) Synergistic infection of two viruses MCMV and SCMV increases the accumulations of both MCMV and MCMV-derived siRNAs in maize. Sci Rep 6. https://doi.org/10.1038/srep20520

Xia ZH, Peng J, Li YQ, Chen L, Li SA, Zhou T, Fan ZF (2014) Characterization of small interfering RNAs derived from Sugarcane mosaic virus in infected maize pPlants by deep sequencing. Plos One 9. https://doi.org/10.1371/journal.pone.0097013

Xie L, Zhang J, Wang Q, Meng C, Hong J, Zhou X (2011) Characterization of *Maize chlorotic mottle virus* associated with Maize lethal necrosis disease in China. Journal of Phytopathology 159:191–193. https://doi.org/10.1111/j.1439-0434.2010.01745.x

Xing Y, Ingvardsen C, Salomon R, Lubberstedt T (2006) Analysis of sugarcane mosaic virus resistance in maize in an isogenic dihybrid crossing scheme and implications for breeding potyvirus-resistant maize hybrids. Genome 49:1274–1282

Xu ML, Melchinger AE, Xia XC, Lübberstedt T (1999) High-resolution mapping of loci conferring resistance to *Sugarcane mosaic virus* in maize using RFLP, SSR, and AFLP markers. Mol Gen Genet 261:574–581

Xu ML, Melchinger AE, Lübberstedt T (2000) Origin of *Scm1* and *Scm2*—two loci conferring resistance to *Sugarcane mosaic virus* (SCMV) in maize. Theor Appl Genet 100:934–941

Yuan L, Dussle CM, Muminovic J, Melchinger AE, Lubberstedt T (2004) Targeted BSA mapping of *Scmv1* and *Scmv2* conferring resistance to SCMV using *PstI/MseI* compared with *EcoRI/MseI* AFLP markers. Plant Breed 123:434–437

Zambrano JL, Francis DM, Redinbaugh MG (2013) Identification of resistance to *Maize rayado fino virus* in maize inbred lines. Plant Dise 97:1418–1423. https://doi.org/10.1094/pdis-01-13-0037-re

Zambrano JL, Jones MW, Brenner E, Francis DM, Tomas A, Redinbaugh MG (2014) Genetic analysis of resistance to six virus diseases in a multiple virus-resistant maize

inbred line. Theor Appl Genet 127:867–880. https://doi.org/10.1007/s00122-014-2263-5

Zhang SH, Li XH, Wang ZH, George ML, Jeffers D, Wang FG, Liu XD, Li MS, Yuan LX (2003) QTL mapping for resistance to SCMV in Chinese maize germplasm. Maydica 48:307–312

Zhang X, Zhang X, Singh J, Li D, Qu F (2012) Temperature-dependent survival of *Turnip crinkle virus*-infected Arabidopsis plants relies on an RNA silencing-based defense that requires DCL2, AGO2, and HEN1. J Virol 86:6847–6854. https://doi.org/10.1128/JVI.00497-12

Zhang Z, Yang F, Na R, Zhang X, Yang S, Gao J, Fan M, Zhao Y, Zhao J (2014) AtROP1 negatively regulates potato resistance to *Phytophthora infestans* via NADPH oxidase-mediated accumulation of H_2O_2. BMC Plant Biol 14:392. https://doi.org/10.1186/s12870-014-0392-2

Zhang ZY, Wang YG, Shen XJ, Li L, Zhou SF, Li WC, Fu FL (2013) RNA interference-mediated resistance to maize dwarf mosaic virus. Plant Cell Tissue Organ Cult 113:571–578. https://doi.org/10.1007/s11240-013-0289-z

Zhao J, Liu Q, Zhang H, Jia Q, Hong Y, Liu Y (2013) The Rubisco small subunit is involved in tobamovirus movement and *Tm-22*-mediated extreme resistance. Plant Physiol 161:374–383

Zhao J, Zhang X, Hong Y, Liu Y (2016) Chloroplast in plant-virus interaction. Front Microbiol 7:1–20. https://doi.org/10.3389/fmicb.2016.01565

Zhou Y, Xu Z, Duan C, Chen Y, Meng Q, Wu J, Hao Z, Wang Z, Li M, Yong H, Zhang D, Zhang S, Weng J, Li X (2016) Dual transcriptome analysis reveals insights into the response to *Rice black-streaked dwarf* virus in maize. J Exp Bot 67:4593–4609. https://doi.org/10.1093/jxb/erw244

Zhu M, Chen YT, Ding XS, Webb SL, Zhou T, Nelson RS, Fan ZF (2014) Maize Elongin C interacts with the viral genome-linked protein, VPg, of *sugarcane mosaic virus* and facilitates virus infection. New Phytologist 203:1291–1304. https://doi.org/10.1111/nph.12890

Genomics of Fungal Disease Resistance

13

Randall J. Wisser and Nick Lauter

Abstract

Fungal diseases are prevalent on maize, for which resistance is controlled by numerous genes where sequence variation more typically gives rise to quantitative rather than qualitative phenotypes. Genomics is facilitating advances in genetics and systems biology while opening the door for convergence between the two. As this is leading to new perspectives about the nature of functionality versus variability during pathogenesis, changes may be afoot in how maize breeders handle the challenge of crop protection.

R. J. Wisser (✉)
Department of Plant and Soil Sciences, University of Delaware, Newark, DE, USA
e-mail: rjw@udel.edu

N. Lauter
Corn Insects and Crop Genetics Research Unit, USDA-ARS, Ames, IA, USA
e-mail: nick.lauter@ars.usda.gov

N. Lauter
Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA, USA

13.1 Introduction

13.1.1 Diseases of Maize: Fungi Reign Supreme

In a compendium of maize diseases, readers will find that fungal pathogens far outnumber other types of attackers (Munkvold and White 2016). As a group, fungal pathogens of maize infect or cause disease on every tissue of a maize plant, resulting in root, stalk and ear rots, kernel molds, foliar blights, rusts, blotches, stripes and spots, and smuts that deform the aforementioned tissues and also the tassel. These pathogens are responsible for substantial losses to maize production worldwide (Oerke et al. 1999; Mueller et al. 2016), and several have been associated with disease outbreaks (e.g., Krauz et al. 1993; Tatum 1971). *Cochliobolus heterostrophus* is infamous for the Southern corn leaf blight (SLB) epidemic of the 1970s (Ullstrup 1972). Fortunately, maize has abundant natural variation that can be—and has been—used as an environmentally sound method of disease control. Genetically engineered resistance to fungal pathogens has not been developed in maize production systems. This is possibly because, at present, the single, major gene engineering techniques currently employed are unlikely to deliver resistance that would withstand the rapid evolution of virulence by fungal pathogens, limiting investments into such engineered resistance. Research on the

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_13

biology of pathogenesis and mechanisms of defense continue to pave the way toward new solutions for disease control, and genomics, considered herein as pertaining to genome-wide investigation, presents new opportunities for progressing toward this aim.

13.1.2 Disease Resistance in Maize: Where is the Qualitative Variation?

Early, key advances toward the understanding of plant disease resistance were made in maize based on genetic analysis of qualitative disease resistance, where essentially all of the segregating phenotypic variation is ascribed to a single gene (Bennetzen et al. 1988; Johal and Briggs 1992). However, a striking feature of maize is the preponderance of quantitative disease resistance (QDR), for which the variation is conditioned by many loci (Wisser et al. 2006). This may seem unexpected to investigators of other plant species, but this comes as little surprise to investigators of maize, where quantitative variation attributed to multigenic architectures is common. Consequently, a range of genomic resources and methods tailored for the dissection of quantitative traits have been created for maize **(see Glowinski and Flint-Garcia, chapter "Germplasm Resources for Mapping Quantitative Traits in Maize" in this book)**, which elevates maize as an excellent model for genomic analysis of disease resistance, particularly QDR. Nevertheless, complexity of the maize genome, both in terms of its size and diversity, poses substantial challenges to unraveling details about the genomic basis of disease resistance.

In agricultural production environments, qualitative resistance is notoriously ephemeral. In contrast, QDR has a reputation of being durable, meaning that resistance is effective over long periods of time and over large areas of cultivation (Johnson 1983). Indeed, durability is one the most appealing attributes of QDR, often invoked as the primary reason for studying it. Understanding the stability of disease resistance has remained a fundamental question since van der Plank crystalized the distinction between concepts of vertical and horizontal forms of resistance, equated to oligogenic (qualitative) and polygenic (quantitative) resistance, respectively (Van Der Plank 1963, 1966). While the nature of durability remains elusive, the multitude of genes underlying resistance, a distinguishing feature of QDR, is likely to be a critical factor, but so too could be distinct host defense mechanisms that confer QDR if they constrain pathogen evolution. Therefore, studying the genomic basis of QDR, which can reveal the number of genes and the mechanisms by which they act to confer resistance, could unlock the key to new, sustainable methods of disease control. For a deeper dive into quantitative and durable resistance, we suggest several review articles (Poland et al. 2009; St Clair 2010; Kou and Wang 2010; Mundt 2014; French et al. 2016; Corwin and Kliebenstein 2017; Yang et al. 2017; Nelson et al. 2018). In this chapter, we focus on genomics research and applications pertaining to fungal disease resistance in maize, including progress toward understanding the biology of qualitative and quantitative forms of resistance. We highlight how the convergence of information on genetic architecture and mechanisms gleaned from systems biology has the potential to provide new perspectives about the evolution and durability of disease resistance.

13.2 Functionality Versus Variability: A Distinction not Unique to Disease Resistance

In biology, which genes control the function of a phenotype and which genes give rise to phenotypic variation are fundamental questions. Genes that underlie function do not necessarily underlie variation, but genes that underlie variation do underlie function. The quest to elucidate one or the other of these questions sets the stage for the types of experimental approaches employed, which in turn affects the knowledge that can be gained. In terms of improving disease resistance in crop plants, elucidating both functionality and

variability can facilitate engineering approaches, but elucidating variability also offers insight into the evolution of resistance and provides information for breeding.

The molecular machinery of the plant defense response to fungal infection or that which is co-opted by pathogens to achieve virulence is comprised of complex networks implicating a large number of functional genes (Tsuda and Somssich 2015), but do all of these genes necessarily contain allelic variation within populations or the species? This is obviously not the case for qualitative resistance, where resistance is mounted by a suite of functional genes that are conditionally responsive to infection based on functional variation at a single gene; however, knowing the totality of functionally variable genes underlying QDR in maize remains an open question and daunting challenge.

For example, using RNAseq to investigate transcriptional reprogramming across each of the genomes of maize lines B73 and Mo17 in response to infection by *C. heterostrophus*, approximately 35% (≈15,000) of all maize genes were significantly responsive to pathogen infection (S. Chudalayandi, N. Lauter et al., unpublished). In contrast, quantitative trait locus (QTL) mapping of functional variation using a B73 × Mo17 recombinant inbred line population identified a maximum of 12 QTL associated with approximately 70% of the genotypic variation in resistance to SLB (Balint-Kurti et al. 2007). Even under the assumption that allelic variation at multiple genes underlies a given QTL effect, this pales in comparison with the number of functionally associated genes identified by RNAseq.

This example highlights a distinction between approaches that dissect functionality from variability. RNAseq is a highly efficient genomic approach for identifying the repertoire of genes involved in the functionality of plant defense, because this can be determined from analysis of as few as one genotype, which also allows many time points and experimental conditions to be surveyed. In contrast, to determine the set of genes underlying variability requires approaches that survey the vast diversity of maize and test for significant allele effects across the maize genome, an enterprise requiring much greater physical space and labor. Enabled by genomic technologies, however, new genetic designs incorporating broader samples of maize germplasm have facilitated discovery of more loci underlying variation in fungal disease resistance (discussed below). The current picture for cultivated maize is that a relatively small fraction of the functionally associated genes explain a large proportion of the variation in resistance to any one fungal disease. Nonetheless, this smaller fraction appears to constitute tens if not hundreds of genes of diverse functions, presenting the dual challenge of mining numerous resistance alleles across many loci from within the complex genome of maize.

13.3 Experimental Platforms Open the Door for Convergence of Genetics and Systems Biology

13.3.1 Bread and Butter: Dissecting the Genetic Architecture of Disease Resistance

For the last several decades, genome-wide mapping studies have dominated the genomics literature on fungal disease resistance in maize. Synthesis studies flanking the decade of 1995–2005 showed a fourfold increase in the number of QTL mapping reports on resistance to fungal pathogens (cf. McMullen and Simcox 1995; Wisser et al. 2006). Meta-analysis of these QTL maps revealed that fungal resistance-associated loci are distributed across most of the maize genome, that resistance to a single fungal disease could be conditioned by at least 50 QTL segregating across maize germplasm (e.g., for Northern leaf blight [NLB], as synthesized in Wisser et al. 2006), and that some regions of the maize genome contain clusters of QTL for resistance to different fungal (and other) diseases (i.e., suggesting that genes may confer multiple disease resistance). However, these QTL maps, constructed prior to the readily accessible high-throughput sequencing technologies available

today, were typically based on standard biparental populations with no more than 250 progeny, which limited the number of QTL that could be detected and biased the estimates of their effects (Beavis 1997), let alone lacked the resolution to confidently implicate specific genes underlying QDR and whether QTL clusters are due to linkage or pleiotropy.

Following the release of the first reference sequence of the maize genome (Schnable et al. 2009), which itself was made possible by advances in sequencing technologies, designs used to dissect the genetic architecture of QDR (and other traits in maize) began to undergo major shifts, where large population sizes and high-density molecular marker maps have been combined with broader surveys of allelic variation. Most notably in maize, a nested association mapping (NAM) population comprised of 5,000 recombinant inbred lines with millions of markers has been used to perform joint linkage and association mapping of variation in QDR to different fungal diseases (Kump et al. 2011; Poland et al. 2011; Benson et al. 2015). These larger-scale mapping studies have helped to further resolve the genetic architecture of QDR: The wide spectrum of variation in resistance captured by as few as 26 maize lines (parents of the NAM population) is largely explained by a set of alleles drawn from 15 to 30 QTL distributed across the genome with multiple, small haplotype effects. This suggests a potent source of variation for the evolution of disease resistance and breeding for defense to fungal pathogens. For instance, in the simplified case of an architecture of 15 QTL and two haplotype effects at each QTL, there are ≈14E6 possible genotypic configurations that could be generated from a biparental cross (computed for a diploid organism, under the admittedly weak assumption of independence among loci: $C = \prod_i^L \left[\frac{n_i(n_i+1)}{2}\right]$, where C is the number of combinations, L is the number of QTL, and n is the number of haplotype effects). This number of combinations far exceeds the population sizes used to characterize the genetic architecture, such that a miniscule fraction of possible combinations are realized (few, if any, would be replicated) in genomic studies of a given experimental population. Theoretically, such an architecture allows progeny with permutations of the parental alleles to produce the same resistance phenotype as they confer resistance via distinct sets of genes—and by extension distinct sets of mechanisms—adding a layer of complexity to the idea that complexity per se may be a key to durable disease resistance (Nelson et al. 2018).

Based on results thus far, the genetic architecture for QDR to fungal pathogens appears to be similar to that of other quantitative traits in maize. Summarizing 41 genetic architectures in terms of the number of loci and the proportion of genotypic variation explained by those loci showed that the genetic architecture for QDR to NLB and SLB diseases falls within the central tendency of these distributions (Wallace et al. 2014). This seems surprising given that the evolution of host resistance uniquely involves the interaction with pathogens that have the capacity for dynamic evolutionary change. Perhaps the evolutionary process for quantitative variation per se or the history of maize domestication and breeding dominates any distinct genomic signatures arising from the evolution of plant-pathogen interactions. Deeper comparative analysis of genetic architectures, identification of the genes underlying functional variation, and systematic analysis of the trade-offs between resistance and other traits may offer further insight into the biology and evolution of disease resistance in maize and other crop species.

13.3.2 Genomic Technologies Unveil Layered Mechanisms

Advanced proteomic approaches used in a new study by Walley et al. (2018) have provided long-awaited mechanistic insights into the virulence of *Cochliobolus carbonum*, a necrotrophic pathogen that causes Northern leaf spot on maize. The maize *C. carbonum* pathosystem has been studied for more than 50 years, beginning with the identification of the *Hm1* locus, which confers qualitative resistance to *C. carbonum*

race 1 (Scheffer et al. 1967). This elegant host–pathogen interaction system permitted the chemical identification of *Helminthosporium carbonum* toxin (HCT), a host-selective cyclic tetrapeptide toxin secreted by *C. carbonum* race 1 (Gross et al. 1982; Walton et al. 1982), as well as the genetic characterization of the large and complex *TOX2* locus that produces it (Panaccione et al. 1992; Scott-Craig et al. 1992; Walton et al. 1994). On the host side, *Hm1* encodes a carbonyl reductase (Johal and Briggs 1992) that can biologically inactivate HCT (Meeley and Walton 1991). It was later shown that HCT acts as a histone deacetylase inhibitor (Brosch et al. 1995), establishing the hypothesis that HCT causes infection in *hm1/hm1* maize plants by interfering with reversible histone acetylation. Coupling acetyl-lysine immunopurification with mass spectroscopy (Choudhary et al. 2009) to globally profile protein abundance and lysine acetylation, Walley et al. (2018) demonstrated that hyperacetylation of host proteins caused by HCT-interference with maize histone deacetylase function results in susceptibility of maize to *C. carbonum* race 1, revealing a "by proxy" virulence mechanism impairing the maize defense response and cellular function.

To obtain specific knowledge of how virulence is achieved by HCT, Walley and colleagues (2018) treated *hm1/hm1* maize plants with HCT + and HCT-strains of *C. carbonum* as well as with exogenously applied HCT and mock HCT solutions and then quantified abundances for non-enriched and lysine acetylated proteins. Abundance was quantified for 3,636 proteins, of which 171 and 116, respectively, increased or decreased in response to both HCT+ treatments. Among the ~5% of the proteome that increased abundance in response to HCT+ treatments, a strong enrichment for indole/tryptophan biosynthetic enzymes was observed, leading to two non-mutually exclusive mechanistic hypotheses for how HCT could establish virulence: (1) Increased auxin levels are associated with elevated indole/tryptophan production and are known to promote susceptibility in several pathosystems; (2) indole is the precursor for the benzoxazinoid class of phytoalexins, which were biochemically shown to be elevated, thereby constituting an inappropriate defense response associated with susceptibility. Lysine acetylation was observed for 2,791 sites corresponding with 912 proteins. Of these, 65 and 9, respectively, increased or decreased in response to both HCT + treatments, with hyperacetylation observed for both histone and non-histone proteins, which is consistent with HCT acting as a deacetylase inhibitor. The majority of hyperacetylated proteins are associated with various components of transcription, including several transcription factors known to be associated with plant immunity in other systems. Notably, the *Zm*MYC2 protein is hyperacetylated, and its resultant inactivity likely causes the aforementioned elevated levels of benzoxazinoid phytoalexins. While further experimentation will be required to determine the relative importance of each of these potential virulence mechanisms, keen insights into the mechanistic complexities of both pathogen defense and susceptibility avoidance were gained.

13.3.3 Realizing the Power of Genetic Segregation

The availability of maize genome sequence and annotation has also enabled system-level transcriptomic dissection of host–pathogen interactions involving maize. For example, a recent study by Christie et al. (2017) measured disease resistance and global gene expression in a segregating population of biparental recombinant inbred lines (RILs) to elucidate networks of genes involved in defense against *Cercospora zeina*, one of the two causal agents of gray leaf spot (GLS) on maize. The authors identified eight QTL for QDR and showed that one-fifth of the expressed genes interrogated in the study were differentially expressed in accordance across the quantitative disease states of RILs. To interconnect these findings, the authors then integrated phenotypic, co-expression, and expression QTL (eQTL) methods to construct transcriptional networks associated with defense. From the thousands of differentially expressed genes, 42 co-expression network modules were identified.

To test whether or not these co-expression modules were under specific genetic control, module eigengene eQTL analysis was performed (Hansen et al. 2008). Remarkably, two of the loci regulating key defense-gene-associated modules mapped to genomic locations coincident with QTL for QDR, suggesting that the functional polymorphisms at these loci are responsible for large-scale alterations of the defense transcriptome, a phenomenon previously observed in barley-Ug99 stem rust interactions (Moscou et al. 2011). Indeed, genome-wide identification of several dozen trans-eQTL hotspots included four hotspots coincidently located with QTL for QDR, revealing deep and multilayered impacts on the defense transcriptome. Collectively, these four loci affect the expression of 874 genes, which are statistically enriched for a myriad of particular functional annotation categories associated with either defense or susceptibility. In summary, this system approach leveraged the power of genetic segregation to make sense of an otherwise uninterpretable number of gene expression polymorphisms, resulting in elucidation of specific molecular processes affected by disease resistance QTL and producing lists of genes that act at particular levels in the cascade of defense responses. The eventual elucidation of the exact polymorphisms underlying these key resistance loci is likely to greatly enrich our understanding of QDR. Particularly within the specific context of functionality versus variability, such genetical genomic studies represent a powerful tool for elevating our understanding of QDR to a level at which it could be strategically manipulated.

13.3.4 Of Oxylipins and Terpenoids: Pathology Unlocked by Metabolomics

Genome-enabled biochemical studies of maize-fungal interactions have produced an array of breakthrough findings over the past decade, revealing new depths of intricacy during maize-fungal interactions. For this review, we focus on oxygenated polyunsaturated fatty acids and terpenoids as two superclasses of metabolites that illustrate the rich biological complexities of biochemical defenses against pathogenic fungi in maize.

Oxylipins, a category of more than 600 metabolites in plants, are derived through both auto-oxidation and enzymatic oxidation of either membrane-esterified or free cytosolic polyunsaturated fatty acids (Borrego and Kolomiets 2016). Enzymatically, the large arrays of metabolites are produced from 18:2 (linoleic) and 18:3 (linolenic) fatty acid substrates by the sequential action of lipoxygenases, oxide synthases, oxide cyclases, and iterative action of up to many beta-oxidases. For example, jasmonic acid, the best known oxylipin in plants, is produced from linolenic acid through the action of seven enzymes. In B73, 13 genes are known to encode lipoxygenases belonging to three classes based on substrate specificity. Six of these gene products, termed 13-LOXs, act specifically on linolenic acid to begin production of the array of jasmonates (Borrego and Kolomiets 2016). Another five lipoxygenases, termed 9-LOXs, act specifically on linoleic acid to begin production of the array of pathologically important death acids (Christensen et al. 2015). The remaining two genes encode enzymes that can act on both linoleic and linolenic acids (Borrego and Kolomiets 2016). Each level of these pathways involves increasing numbers of enzymes as the diversity of metabolites grows greater.

Two important 9-LOX-derived metabolites in maize necrotroph biology are the death acids (DAs) 10-oxo-11-phytodienoic acid (10-OPDA) and (10-oxo-11-phytoenoic acid (10-OPEA), which were first characterized in maize using leaf tissue infected with *C. heterostrophus*, the causal agent of SLB (Christensen et al. 2015). These DAs and their derivatives display complex roles in maize defense, including acting as cytotoxic agents that kill host cells (a containment mechanism), triggering cascades of defense gene expression, and acting as defensive phytoalexins in response to some pathogenic fungi (Christensen et al. 2014, 2015). Unraveling such complexities will likely require genetic manipulations of the genes encoding the specific

enzymes that affect levels of specific oxylipin metabolites. An example of this approach has been executed to study derivatives of 12-OPDA produced by 12-oxo-phytodienoate reductase genes *Opr7* and *Opr8*, revealing complex roles in both development and defense for these jasmonate-class oxylipins (Yan et al. 2012).

Another particularly interesting aspect of oxylipin biology is that they occur with equal complexity and ubiquity in fungi and animals (e.g., prostaglandins and leukotrienes). Moreover, they are known agents of cross-kingdom communication during biotic interactions and are particularly important in maize-fungal interactions (reviewed by Christensen and Kolomiets 2011). It appears that these ancient and ever-present oxylipin-mediated interactions have produced a complex interactome over evolutionary time, involving host-ligand mimicry, co-option, and deceit. As such, targeted manipulation of these metabolites for crop improvement may require extensive understanding of their roles across an array of biotic interactions.

Terpenoids are the largest category of specialized metabolites produced by plants, with more than 25,000 analytes known across the plant kingdom, many of which are elicited during pathogen and herbivore defense (Schmelz et al. 2014). Especially important for fungal defense are the nonvolatile sesquiterpenoids and diterpenoids that are, respectively, derived from farnesyl diphosphate (FDP) and geranylgeranyl diphosphate (GGDP), which each arise via multiple enzymatic steps from isopentenyl diphosphate. The biosynthesis of diverse sesquiterpenoid zealexins occurs through the sequential action of terpene synthase and cytochrome p450 mono-oxygenase enzymes and is induced by inoculation of many different fungal pathogens of maize, with the notable exception of *Colletotrichum graminicola*, the causal agent of anthracnose stalk rot (Huffaker et al. 2011). The biosynthesis of diterpenoid phytoalexins begins with the sequential action of both a type I and a type II diterpene synthase to give rise to various precursors, which are then enzymatically modified to produce diverse arrays of kauralexin and dolabralexin defense metabolites (Schmelz et al. 2011, 2014; Mafu et al. 2018). In contrast with zealexins, kauralexins effectively suppress growth of *C. graminicola*, as well as other diverse fungal pathogens, highlighting the selective nature of biochemical efficacy in defense (Meyer et al. 2017; Christensen et al. 2018). A challenge ahead is to identify the specific enzymes that act in the final step to produce the individual zealexin, kauralexin, and dolabralexin metabolites, as this will enable focused studies on elicitation and specific activities that are relevant for manipulating their presence and abundance for crop protection.

13.4 Genomic Selection and Allele Mining for Fungal Disease Resistance

Genomic selection is a breeding procedure that uses an index of estimated allele effects (determined from a training population) for every marker across the genome in order to predict phenotypic values for unobserved individuals based solely on their genotypic makeup (Meuwissen et al. 2001). Given the multigenic architecture of most fungal disease resistance in maize, the environmental dependence of disease pressure during screening, and the difficulty of phenotyping some fungal diseases, genomic selection can be considered as a strategy to develop populations enriched with resistance alleles and to develop resistant varieties (Poland and Rutkoski 2016). While the operational principles for genomic selection normally transcend the characteristics to which the method is applied, the potential strain-specific nature of QDR may challenge this. As such, genomic selection for disease resistance may uniquely benefit from integrating information on genetic diversity about the local pathogen populations to which resistance is bred. For many agronomically relevant traits, genomic selection has begun to displace conventional breeding methodology and literature on the subject grows larger. However, examples of genomic selection for fungal disease resistance in maize are missing: We are aware of no study that has yet demonstrated

genomic selection for disease resistance in maize, and there is a paucity of reports on genomic prediction which sets expectations for how well genomic selection would perform.

There are four studies, with promising findings overall, on genomic prediction (not selection) for fungal disease resistance in maize. Technow et al. (2013) estimated prediction accuracies of ≈0.7 for resistance to NLB (caused by *Setosphaeria turcica*) in the context of breeding of maize hybrids. A prediction accuracy of approximately ≈0.7 was also estimated for resistance to ear rot caused by *Stenocarpella maydis*; we computed prediction accuracy ($r_{y\hat{y}} = r_{g\hat{g}}/h$) using the prediction ability ($r_{g\hat{g}}$) and the square root of mean heritability (h) reported for the proportion of rotten kernels by dos Santos et al. (2016). For the mycotoxin deoxynivalenol, produced in maize kernels by *Fusarium graminearum*, Han et al. (2016) reported family-dependent and model-dependent maximum prediction accuracies ranging from ≈0.25 to ≈0.55. Lastly, in an ongoing investigation of genomic selection for Fusarium ear rot and fumonisin content (caused by *Fusarium verticillioides*), prediction accuracies in a breeding population were estimated to be ≈0.6 and ≈0.7, respectively (T. Marino, J. Holland, pers. comm.). These results suggest genomic selection is a viable option for fungal disease resistance improvement in maize. With only these few studies published thus far, there is ample opportunity for further investigation into genomic selection, including extensions that integrate evolutionary principles of host–pathogen dynamics. Moreover, given that genotype data on extant maize germplasm continues to become increasingly available, preexisting genomic prediction models could be useful for mining maize germplasm and alleles (Yu et al. 2016).

As a complement to genomic selection, infusions of critical alleles may periodically be required, particularly in response to instances of widespread poor performance due to disease susceptibility. However, knowledge of the causal genes and the spectrum of functional allelic variation underlying resistance will be required to enable this approach. Even as allele mining capabilities increase, genome editing technologies using the CRISPR–Cas system (Barrangou and Horvath 2017) are poised to become a key approach to stacking favorable alleles for disease resistance, or replacing unfavorable alleles with those that are at least neutral. Beyond mechanistic characterization of causal genes, there may also exist the need for highly efficient screening systems that allow variant effects of natural or synthetic alleles to be functionally evaluated across multiple contexts.

13.5 Conclusions and Frontiers

Fungal diseases pose continued threats to the quality, affordability, and availability of maize products in food markets and to price stability in feed, fuel, and processing markets. As maize product markets change and diversify and as temperate climates experience warmer winters, disease control will likely not get easier. However, both the maize genome and maize genomics are being leveraged to tackle challenges to the provision of food, feed, fuel, and fiber for a growing population. The maize genome is remarkably well-equipped, versatile, and pliable, and our understanding of essential disease resistance mechanisms is improving at the DNA, RNA, protein, and metabolite levels, as highlighted by examples of breakthroughs using quantitative genetic, genetical genomic, proteomic, and metabolomic approaches. Nevertheless, increased emphasis on elucidation of QDR mechanisms is warranted for several reasons. First, crop improvement via engineering requires this information for success—we must know which genes contribute key functionality and variability and how the most valuable alleles should act in diverse contexts. Next, stress biology research is a key path toward a more thorough functional characterization of the maize genome, an obvious goal for humankind in view of the fact that maize is the most productive and most widely grown crop in our world. At present, few (we estimate no more than 10%) of the

approximately 40,000 maize genes have empirical evidence describing their function—this is too large and dark of a box from which to draw the tools we need to make elegant manipulations of maize, let alone to understand the trade-offs that are inherent for so many types of alterations.

It is difficult to assess, and therefore to predict, which research approaches are the most important to pursue in deepening our understanding of maize-fungal interaction biology. In many ways, this scale may be tipped by the relative weights of functionality and variability on a per gene basis, making the right choice an impossibility for an individual investigator, regardless of their calculus. That said, an apparent similarity in the genetic architecture for QDR and other quantitative traits has clearly emerged, suggesting that continued mining of the variable products of evolution remains a worthwhile pursuit in support of both conventional allele usage and synthetic allele deployment. This will likely remain true, regardless of which ancillary approaches emerge as essential.

As we make progress in understanding fungal disease resistance at accelerating rates, how should our expectations for agronomic outcomes change? Perhaps a good answer at the moment rests in a series of additional questions. To what extent is disease resistance knowledge having an impact on maize production systems? Will new methods for developing resistant varieties be more profitable for seed companies and for growers? What are the trade-offs and limits to crop protection at different localities and scales? How are these efforts helping to alleviate global food and health insecurity? While these questions offer guidance, we expect that progress achieved in the recent past is a good predictor of future research accomplishments regarding disease resistance, forecasting success in the defense of maize.

Acknowledgements This work was made possible by the US NSF Plant Genome Research Program IOS-1127076 and the US Department of Agriculture—Agricultural Research Service.

References

Balint-Kurti PJ, Zwonitzer JC, Wisser RJ et al (2007) Precise mapping of quantitative trait loci for resistance to southern leaf blight, caused by *Cochliobolus heterostrophus* race O, and flowering time using advanced intercross maize lines. Genetics 176:645–657

Barrangou R, Horvath P (2017) A decade of discovery: CRISPR functions and applications. Nat Microbiol 2:1–9

Beavis WD (1997) QTL analysis, power, precision, and accuracy. In: Paterson A (ed) Molecular dissection of complex traits. CRC Press, New York, pp 145–162

Bennetzen JL, Qin M-M, Ingels S, Ellingboe AH (1988) Allele-specific and mutator-associated instability at the *Rp1* disease-resistance locus of maize. Nature 332:369–370

Benson JM, Poland JA, Benson BM et al (2015) Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. PLoS Genet 11:1–23

Borrego E, Kolomiets M (2016) Synthesis and functions of jasmonates in maize. Plants 5(4):E41

Brosch G, Ransom R, Lechner T et al (1995) Inhibition of maize histone deacetylases by HC toxin, the host-selective toxin of *Cochliobolus carbonum*. Plant Cell 7:1941–1950

Choudhary C, Kumar C, Gnad F et al (2009) Lysine acetylation targets protein complexes and co-regulated major cellular functions. Science 325:834–840

Christensen SA, Huffaker A, Kaplan F et al (2015) Maize death acids, 9-lipoxygenase–derived cyclopente(a) nones, display activity as cytotoxic phytoalexins and transcriptional mediators. Proc Natl Acad Sci 112:11407–11412

Christensen SA, Kolomiets MV (2011) The lipid language of plant-fungal interactions. Fungal Genet Biol 48:4–14

Christensen SA, Nemchenko A, Park Y-S et al (2014) The novel monocot-specific 9-lipoxygenase *Zm*LOX12 is required to mount an effective jasmonate-mediated defense against *Fusarium verticillioides* in maize. Mol Plant-Microbe Interact 27:1263–1276

Christensen SA, Sims J, Vaughan MM et al (2018) Commercial hybrids and mutant genotypes reveal complex protective roles for inducible terpenoid defenses in maize. J Exp Bot 69:1693–1705

Christie N, Myburg AA, Joubert F et al (2017) Systems genetics reveals a transcriptional network associated with susceptibility in the maize–grey leaf spot pathosystem. Plant J 89:746–763

Corwin JA, Kliebenstein DJ (2017) Quantitative resistance: more than just perception of a pathogen. Plant Cell 29:655–665

dos Santos JPR, Pires LPM, de Castro Vasconcellos RC et al (2016) Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. BMC Genet 17:1–10

French E, Kim BS, Iyer-Pascuzzi AS (2016) Mechanisms of quantitative disease resistance in plants. Semin Cell Dev Biol 56:201–208

Gross ML, McCrery D, Crow F et al (1982) The structure of the toxin from helminthosporium carbonum. Tetrahedron Lett 23:5381–5384

Han S, Utz HF, Liu W et al (2016) Choice of models for QTL mapping with multiple families and design of the training set for prediction of Fusarium resistance traits in maize. Theor Appl Genet 129:431–444

Hansen BG, Halkier BA, Kliebenstein DJ (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. Trends Plant Sci 13:72–77

Huffaker A, Kaplan F, Vaughan MM et al (2011) Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. Plant Physiol 156:2082–2097

Johal GS, Briggs SP (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. Science 258:985–987

Johnson R (1983) Genetic background of durable resistance. Durable resistance in crops. Springer, New York, pp 5–26

Kou Y, Wang S (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. Curr Opin Plant Biol 13:181–185

Krauz JP, Fredericksen RA, Rodrigues-Ballesteros OR (1993) Epidemic of northern corn leaf blight in Texas in 1992. Plant Disease 77:1063

Kump KL, Bradbury PJ, Wisser RJ et al (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–169

Mafu S, Ding Y, Murphy KM et al (2018) Discovery, biosynthesis and stress-related accumulation of dolabradiene-derived defenses in maize. Plant Physiol 176:2677–2690

McMullen M, Simcox KD (1995) Genomic organization of disease and insect resistance genes in maize. Mol Plant-Microbe Interact 8:811–815

Meeley RB, Walton JD (1991) Enzymatic detoxification of HC-toxin, the host-selective cyclic peptide from Cochliobolus carbonum. Plant Physiol 97:1080–1086

Meyer J, Berger DK, Christensen SA, Murray SL (2017) RNA-Seq analysis of resistant and susceptible sub-tropical maize lines reveals a role for kauralexins in resistance to grey leaf spot disease, caused by *Cercospora zeina*. BMC Plant Biol 17:1–20

Moscou MJ, Lauter N, Steffenson B, Wise RP (2011) Quantitative and qualitative stem rust resistance factors in barley are associated with transcriptional suppression of defense regulons. PLoS Genet 7(7): e1002208

Mueller DS, Wise KA, Sisson AJ et al (2016) Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Plant Heal Prog 17:211–222

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Mundt CC (2014) Durable resistance: a key to sustainable management of pathogens and pests. Infect Genet Evol 27:446–455

Munkvold GP, White DG (2016) Compendium of corn diseases, 4th edn. APS Press, The American Phytopathological Society, St. Paul

Nelson R, Wiesner-Hanks T, Wisser R, Balint-Kurti P (2018) Navigating complexity to breed disease-resistant crops. Nat Rev Genet 19:21–33

Oerke E-C, Dehne H-W, Schönbeck F et al (1999) Estimated crop losses due to pathogens, animal pests and weeds. Crop production and crop protection. Elsevier, New York, pp 72–741

Panaccione DG, Scott-Craig JS, Pocard JA, Walton JD (1992) A cyclic peptide synthetase gene required for pathogenicity of the fungus *Cochliobolus carbonum* on maize. Proc Natl Acad Sci USA 89:6590–6594

Poland JA, Balint-Kurti PJ, Wisser RJ et al (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci 14:21–29

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci USA 108:6893–8

Poland J, Rutkoski J (2016) Advances and challenges in genomic selection for disease resistance. Annu Rev Phytopathol 54:79–98

Scheffer RP, Nelson RR, Ullstrup AJ (1967) Inheritance of toxin production and pathogenicity in *Cochliobolus carbonum* and *Cochliobolus victoriae*. Phytopathology 57:1288

Schmelz EA, Huffaker A, Sims JW et al (2014) Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. Plant J 79:659–678

Schmelz EA, Kaplan F, Huffaker A et al (2011) Identity, regulation, and activity of inducible diterpenoid phytoalexins in maize. Proc Natl Acad Sci 108:5455–5460

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Scott-Craig JS, Panaccione DG, Pocard JA, Walton JD (1992) The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. J Biol Chem 267:26044–26049

St Clair DA (2010) Quantitative disease resistance and quantitative resistance Loci in breeding. Annu Rev Phytopathol 48:247–68

Tatum LA (1971) The southern corn leaf blight epidemic. Science 171:1113–1116

Technow F, Bürger A, Melchinger AE (2013) Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for

heterotic groups. G3 Genes Genomes Genet 3:197–203

Tsuda K, Somssich IE (2015) Transcriptional networks in plant immunity. New Phytol 206:932–947

Ullstrup AJ (1972) The impacts of the southern corn leaf blight epidemics of 1970–1971. Annu Rev Phytopathol 10:37–50

Van Der Plank JE (1963) Plant diseases: epidemics and control. Academic Press, New York

Van Der Plank JE (1966) Horizontal (polygenic) and vertical (oligogenic) resistance against blight. Am Potato J 43:43–52

Wallace JG, Bradbury PJ, Zhang N et al (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet 10(12): e1004845

Walley JW, Shen Z, McReynolds MR et al (2018) Fungal-induced protein hyperacetylation in maize identified by acetylome profiling. Proc Natl Acad Sci 115:210–215

Walton JD, Akimitsu K, Ahn JH, Pitkin JW (1994) Towards an understanding of the *TOX2* gene of *Cochliobolus carbonum*. In: Kohmoto K, Yoder OC (eds.) Proceedings of the second Tottori University international symposium on host-specific toxin: biosynthesis, receptor and molecular biology, pp 227–237

Walton JD, Earle ED, Gibson BW (1982) Purification and structure of the host-specific toxin from *Helminthosporium carbonum* race 1. Biochem Biophys Res Commun 107:785–794

Wisser RJ, Balint-Kurti PJ, Nelson RJ (2006) The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96:120–129

Yan Y, Christensen S, Isakeit T et al (2012) Disruption of *OPR7* and *OPR8* reveals the versatile functions of jasmonic acid in maize development and defense. Plant Cell 24:1420–1436

Yang Q, Balint-Kurti P, Xu M (2017) Quantitative disease resistance: dissection and adoption in maize. Mol Plant 10:402–413

Yu X, Li X, Guo T et al (2016) Genomic prediction contributing to a promising global strategy to turbocharge gene banks. Nat Plants 2:1–7

Endophytes: The Other Maize Genome

14

Jason G. Wallace and Georgiana May

Abstract

Endophytes are microorganisms that live inside plants without causing disease. Maize endophytes collectively encode roughly twenty times as many genes as maize itself, giving the plant access to incredible genetic diversity. They can affect their host plant by altering growth, nutrient acquisition, disease resistance, insect resistance, and abiotic stress tolerance. Despite the richness of these communities, the rules that govern their assembly and their functions within the maize plant are complex and poorly understood. We outline what is known about maize endophytes, including which organisms are known to live inside maize, how they are transmitted, what genomic functions they encode, what effects they have on their host, and how they interact with each other and the maize plant. Many questions still remain about maize endophytes, including what makes a healthy endophyte community, how that community is assembled and develops, and how endophytes can be harnessed to improve agriculture.

14.1 Introduction

One of the great surprises of the genomics era was how few genes most genomes encode. For example, sequencing reduced the number of estimated genes in the human genome from 100,000 (Goodfellow 1995; Pertea and Salzberg 2010) to 30–40,000 (Lander et al. 2001; Venter et al. 2001) and now a mere 20–25,000 (ENCODE Project Consortium 2004); some authors argue the true number may be smaller still (Pertea and Salzberg 2010). The maize genome went through a similar (albeit less extreme) process, with estimates of up to 59,000 genes (Messing et al. 2004) dropping to the current estimate of about 40,000 genes (Jiao et al. 2017), comfortably in the middle range for plant genomes (Michael and Jackson 2013). Subsequently, the hunt to account for the functional complexity of eukaryotic genomes has turned toward other, more dynamic sources.

These raw gene counts do not reflect the whole genomic complexity of an organism because alternative splicing, RNA editing, noncoding RNA transcripts, and a suite of mechanisms to regulate expression can dramatically increase the functional complexity of an organism (Pray 2008; Syed et al. 2012; Takenaka et al. 2013; Liu et al. 2015a).

J. G. Wallace (✉)
Department of Crop & Soil Sciences, University of Georgia, 111 Riverbend Road, 30602 Athens, Georgia
e-mail: jason.wallace@uga.edu

G. May
Department of Ecology, Evolution and Behaviour, University of Minnesota Twin Cities, 140 Gortner Ave, St. Paul, MN 55108, USA

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_14

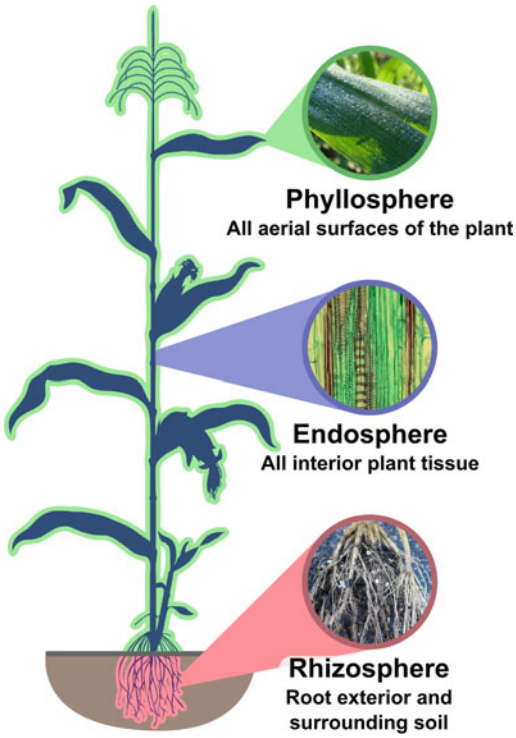

Fig. 14.1 Maize microbiome compartments. The maize microbiome can be roughly divided into three sections. The *phyllosphere* includes all aerial surfaces of the plant (stalk, leaves, etc.). The *rhizosphere* includes the root surfaces (sometimes treated separately as the *rhizoplane*) and the surrounding soil. The *endosphere* includes all interior plant tissues. Maize endophytes, by definition, live in the endosphere, although most are also capable of living elsewhere. (Maize stem cross section by Josef Reischig, CC-SA-BY)

The most surprising source of complexity, however, has turned out to not even be part of the organism itself. We now know that microbes living on and inside an organism can dramatically increase the genomic functions available to it. These microbes are collectively called an organism's *microbiome*. Although some members of a microbiome are just commensals who are occupying a niche, others have important interactions with their plant or animal host. Understanding and exploiting these interactions will likely be an important component of twenty-first-century biology.

In this chapter, we focus on the microbiome of maize. Specifically, we focus on maize endophytes, meaning the microbes that live inside the plant's own tissue (Fig. 14.1). We frame these endophytes as the *other* maize genome due to their intimate association with maize, the range of functional and biochemical repertoires they provide, and the impacts they can have on their plant host. Some authors refer to a plant and its microbiome as a *holobiont* with an associated *hologenome*. We choose to avoid those terms, however, because they imply much stricter vertical inheritance and coevolution than usually occurs with maize endophytes (Moran and Sloan 2015). Instead, we prefer to think of maize endophytes as providing an *extended phenotype*, where the phenotype of the maize plant results from the combined effects of the plant genome, various endophyte genomes, the environment, and the complex interactions among them.

14.2 Endophyte Lifestyles

14.2.1 Known Maize Endophytes

The definition of what constitutes an endophyte is a little fuzzy. It originally meant pathogenic fungi in leaves (Bary 1866), but more recently has been used to include any microorganism that lives inside plants without causing disease (see Proença et al. 2017 for a brief history of the term). Some authors (e.g., Hardoim et al. 2015) have recently proposed using it to include any organisms that can live inside a plant at all, including active pathogens. For the purposes of this review, we keep to the more traditional definition of endophytes being organisms that live within plant tissues without causing apparent disease symptoms (Petrini 1991). While this definition obviously includes commensal and beneficial organisms, it can also include latent pathogens or ones whose detrimental effects on plant health are too subtle to be classified as a disease.

Maize endophytes span a large range of both bacteria and fungi, although most belong to relatively few phyla. Of the ~300 maize-associated endophytes (mostly bacteria) that Hardoim et al. (2015) compiled from the literature, over 90% belonged to either Firmicutes or Proteobacteria,

with the genus *Bacillus* by far the most common, followed by *Burkholderia*, *Enterobacteria*, and *Paenibacillus*. Our own review of the literature (Table 14.1) finds a similar pattern, with most bacterial endophytes in either the Firmicutes or the Proteobacteria. Within these phyla, the most common genera are *Bacillus*, *Paenibacillus*, and *Staphylococcus* (Firmicutes); *Rhizobium/Agrobacterium* and *Sphingomonas* (Alphaproteobacteria); *Burkholderia* (Betaproteobacteria); and *Enterobacter*, *Klebsiella*, *Pantoea*, and *Pseudomonas* (Gammaproteobacteria).

Most fungal endophytes are members of the Ascomycota in the classes Dothideomycetes and Sordariomycetes (Table 14.2). Not surprisingly, the most common genera include ones with known plant pathogens, such as *Alternaria*, *Fusarium*, and *Acremonium*, and known beneficial organisms such as *Trichoderma*. Archaea are only rarely found as maize endophytes (Table 14.3) (Chelius and Triplett 2001; da Silva et al. 2014); those found to date belong to the Thaumarchaeota and Euryarchaeota. Since most authors have not found archaeal endophytes despite using techniques that can include them, archaea probably do not play a significant role in the maize endosphere.

14.2.2 Acquisition and Inheritance

Maize endophytes can be either acquired from the environment (horizontal transmission) or inherited via seed (vertical transmission). The relative importance of these two transmission modes has not been rigorously investigated for maize endophytes, so the normal inheritance pattern is unclear for most of them. Some reports indicate that environment is the major source for most maize endophytes (McInroy and Kloepper 1995a; Bokati et al. 2016), while others show seed transmission as being more important (Johnston-Monje and Raizada 2011; Johnston-Monje et al. 2016). Similarly, the patchy distributions and small infection loci of fungal endophytes on leaves would imply horizontal transmission but do not completely rule out vertical transmission via seeds (Pan and May 2009). This same issue of environment versus inheritance has been raised for endophytes in other systems (e.g., Compant et al. 2010; Kroll et al. 2017); since the various investigations involve different techniques (culturing versus sequencing) and different taxa (e.g., bacteria versus fungi), it is hard to say for certain which is more correct. Given the diversity of plant microbiomes, many modes of transmission are probably present, with the exact balance depending both on the organism in question and on the local environment.

Seed transmission has been observed for several maize endophytes (McInroy and Kloepper 1995a; Johnston-Monje and Raizada 2011; Liu et al. 2012b, 2015b; Johnston-Monje et al. 2014, 2016). In theory, this means that some endophytes could become obligate symbionts with strongly vertical transmission, similar to the *Epichloë* endophytes of tall fescue (Siegel et al. 1984; Schardl 2001). To our knowledge, however, no one has identified an obligate dependence of an endophyte on maize specifically. The closest we know of are the endomycorrhizal fungi (Glomeromycota), which are obligately dependent on their plant hosts. These hosts can include maize, but also many other plants (Helgason and Fitter 2009).

Regarding acquisition from the environment, many endophytes (especially bacteria) are known to enter through the roots (McInroy and Kloepper 1995a; Lamb et al. 1996; Zinniel et al. 2002; Roncato-Maccari et al. 2003; Liu et al. 2006; Monteiro et al. 2008; Prischl et al. 2012; Johnston-Monje et al. 2014; Bokati et al. 2016) Bacteria usually enter via root cracks—small disruptions in the epidermis—near lateral root growth (Roncato-Maccari et al. 2003; Liu et al. 2006; Monteiro et al. 2008). Some endophytes can enter the aerial parts of plants via stomata (Roos and Hattingh 1983) or possibly wounds (Mano and Morisaki 2008), but there are a few reports of this in maize (Lamb et al. 1996). There are no confirmed reports of maize specifically attracting endophytes, such as happens with strigolactone signaling for arbuscular mycorrhizae (Besserer et al. 2006), although the possibility certainly exists. It may even

Table 14.1 Known maize endophytes (bacteria)

Phylum	Class	Genus[a]	References
Acidobacteria	Solibacteres	*Bryobacter*	Mashiane et al. (2017)
Actinobacteria	Actinobacteria	*Agromyces*	Prischl et al. (2012)
Actinobacteria	Actinobacteria	*Arthrobacter*	Johnston-Monje and Raizada (2011), Liu et al. (2012b), McInroy and Kloepper (1995b), Prischl et al. (2012)
Actinobacteria	Actinobacteria	*Brachybacterium*	(Mashiane et al. 2017)
Actinobacteria	Actinobacteria	*Cellulomonas*	Johnston-Monje and Raizada (2011), Orole and Adejumo (2011), Zinniel et al. (2002)
Actinobacteria	Actinobacteria	*Dietzia*	Mashiane et al. (2017)
Actinobacteria	Actinobacteria	*Frigoribacterium*	Chelius and Triplett (2001), Rijavec et al. (2007)
Actinobacteria	Actinobacteria	*Geodermatophilus*	Mashiane et al. (2017)
Actinobacteria	Actinobacteria	*Lentzea*	Liu et al. (2012a)
Actinobacteria	Actinobacteria	*Leucobacter*	Mashiane et al. (2017)
Actinobacteria	Actinobacteria	*Marmoricola*	Mashiane et al. (2017)
Actinobacteria	Actinobacteria	*Microbacterium*	Chelius and Triplett (2001), Gao et al. (2017), Orole and Adejumo (2011), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), McInroy and Kloepper (1995b), Prischl et al. (2012), Rijavec et al. (2007), Zinniel et al. (2002)
Actinobacteria	Actinobacteria	*Microbispora*	de Araujo et al. (2000)
Actinobacteria	Actinobacteria	*Micrococcus*	Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Mashiane et al. (2017), McInroy and Kloepper (1995b), Orole and Adejumo (2011), Zinniel et al. (2002)
Actinobacteria	Actinobacteria	*Nesterenkonia*	Liu et al. (2017)
Actinobacteria	Actinobacteria	*Nocardia*	Liu et al. (2012a)
Actinobacteria	Actinobacteria	*Nocardioides*	Mashiane et al. (2017), Prischl et al. (2012)
Actinobacteria	Actinobacteria	*Plantibacter*	Prischl et al. (2012)
Actinobacteria	Actinobacteria	*Propionibacterium*	Liu et al. (2012a), Mashiane et al. (2017)
Actinobacteria	Actinobacteria	*Propioniciclava*	Liu et al. (2012a)
Actinobacteria	Actinobacteria	*Pseudonocardia*	Liu et al. (2012a)
Actinobacteria	Actinobacteria	*Rathayibacter*	Mashiane et al. (2017)
Actinobacteria	Actinobacteria	*Rothia*	Liu et al. (2012b), Zinniel et al. (2002)
Actinobacteria	Actinobacteria	*Streptomyces*	Chelius and Triplett (2001), de Araujo et al. (2000), Johnston-Monje and Raizada (2011)
Actinobacteria	Actinobacteria	*Thermomonospora*	Mashiane et al. (2017)
Actinobacteria	Actinomycetales	*Amycolatopsis*	Chelius and Triplett (2001)
Actinobacteria	Actinomycetales	*Brevibacterium*	Liu et al. (2012b), Mashiane et al. (2017)
Actinobacteria	Actinomycetales	*Clavibacter*	Gao et al. (2004), McInroy and Kloepper (1995b), Zinniel et al. (2002)

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
Actinobacteria	Actinomycetales	*Corynebacterium*	Bodhankar et al. (2017), Liu et al. (2012a, b), Mashiane et al. (2017), Zinniel et al. (2002)
Actinobacteria	Actinomycetales	*Curtobacterium*	Johnston-Monje et al. (2014), Mashiane et al. (2017), McInroy and Kloepper (1995b), Prischl et al. (2012)
Actinobacteria	Actinomycetales	*Kocuria*	Chelius and Triplett (2001), Liu et al. (2012a, 2017), Prischl et al. (2012)
Actinobacteria	Actinomycetales	*Kytococcus*	Johnston-Monje et al. (2014)
Actinobacteria	Actinomycetales	*Leifsonia*	Chelius and Triplett (2001)
Actinobacteria	Actinomycetales	*Micromonospora*	Chelius and Triplett (2001), Mashiane et al. (2017)
Actinobacteria	Actinomycetales	*Mycobacterium*	Johnston-Monje et al. (2014)
Actinobacteria	Actinomycetales	*Rhodococcus*	Johnston-Monje and Raizada (2011), Mashiane et al. (2017)
Actinobacteria	Actinomycetales	*Streptosporangium*	de Araujo et al. (2000)
Bacteroidetes	Chitinophagia	*Sediminibacterium*	Chelius and Triplett (2001), Johnston-Monje and Raizada (2011)
Bacteroidetes	Cytophagia	*Dyadobacter*	Prischl et al. (2012)
Bacteroidetes	Cytophagia	*Flexibacter*	Johnston-Monje et al. (2014)
Bacteroidetes	Cytophagia	*Hymenobacter*	Mashiane et al. (2017)
Bacteroidetes	Cytophagia	*Siphonobacter*	Mashiane et al. (2017)
Bacteroidetes	Flavobacteriia	*Chryseobacterium*	Arruda et al. (2013), Chelius and Triplett (2001), da Silva et al. (2014), Liu et al. (2012a, b), Mashiane et al. (2017), Prischl et al. (2012)
Bacteroidetes	Flavobacteriia	*Cloacibacterium*	Mashiane et al. (2017)
Bacteroidetes	Flavobacteriia	*Elizabethkingia*	Mashiane et al. (2017)
Bacteroidetes	Flavobacteriia	*Empedobacter*	Mashiane et al. (2017)
Bacteroidetes	Flavobacteriia	*Flavobacterium*	Arruda et al. (2013), da Silva et al. (2014), Liu et al. (2012a, b), Prischl et al. (2012)
Bacteroidetes	Sphingobacteriia	*Chitinophaga*	(Prischl et al. 2012)
Bacteroidetes	Sphingobacteriia	*Flavisolibacter*	Liu et al. (2012a, b)
Bacteroidetes	Sphingobacteriia	*Mucilaginibacter*	Chelius and Triplett (2001)
Bacteroidetes	Sphingobacteriia	*Niastella*	Chelius and Triplett (2001)
Bacteroidetes	Sphingobacteriia	*Nubsella*	Chelius and Triplett (2001)
Bacteroidetes	Sphingobacteriia	*Pedobacter*	Johnston-Monje et al. (2014), Prischl et al. (2012)
Bacteroidetes	Chitinophagia	*Sediminibacterium*	Chelius and Triplett (2001), Johnston-Monje and Raizada (2011)
Bacteroidetes	Sphingobacteriia	*Sphingobacterium*	Liu et al. (2012a), Mashiane et al. (2017)
Chloroflexi	Chloroflexia	*Chloronema*	Mashiane et al. (2017)
Chloroflexi	Chloroflexia	*Roseiflexus*	Mashiane et al. (2017)

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
Deinococcus–Thermus	Deinococci	*Deinococcus*	Johnston-Monje and Raizada (2011), Mashiane et al. (2017)
Firmicutes	Bacilli	*Aeribacillus*	Liu et al. (2017)
Firmicutes	Bacilli	*Aerococcus*	Liu et al. (2012a)
Firmicutes	Bacilli	*Alicyclobacillus*	Mashiane et al. (2017)
Firmicutes	Bacilli	*Bacillus*	Arruda et al. (2013), Bodhankar et al. (2017), Chelius and Triplett (2001), Figueiredo et al. (2009), Gao et al. (2004), Gond et al. (2015), Hinton and Bacon (1995), Ikeda et al. (2013), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b, 2017), Mashiane et al. (2017), Matsumura et al. (2015), McInroy and Kloepper (1995b), Orole and Adejumo (2011), Pereira et al. (2011), Prischl et al. (2012), Riggs et al. (2001), Rijavec et al. (2007), Szilagyi-Zecchin et al. (2014), Zinniel et al. (2002)
Firmicutes	Bacilli	*Brevibacillus*	Johnston-Monje and Raizada (2011)
Firmicutes	Bacilli	*Cohnella*	Johnston-Monje et al. (2014)
Firmicutes	Bacilli	*Desemzia*	Liu et al. (2012a)
Firmicutes	Bacilli	*Enterococcus*	Johnston-Monje and Raizada (2011), Liu et al. (2017)
Firmicutes	Bacilli	*Kurthia*	Orole and Adejumo (2011)
Firmicutes	Bacilli	*Lactobacillus*	Liu et al. (2012a) Mashiane et al. (2017)
Firmicutes	Bacilli	*Lactococcus*	da Silva et al. (2014), Mashiane et al. (2017)
Firmicutes	Bacilli	*Leuconostoc*	Liu et al. (2017), Mashiane et al. (2017)
Firmicutes	Bacilli	*Oxalophagus*	Liu et al. (2012a)
Firmicutes	Bacilli	*Paenibacillus*	Arruda et al. (2013), Chelius and Triplett (2001), da Silva et al. (2014), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b, 2015b, 2017), Prischl et al. (2012), Rijavec et al. (2007), Roesch et al. (2008)
Firmicutes	Bacilli	*Pediococcus*	Mashiane et al. (2017), Orole and Adejumo (2011)
Firmicutes	Bacilli	*Planomicrobium*	Liu et al. (2012b)
Firmicutes	Bacilli	*Sediminibacillus*	Liu et al. (2012a)
Firmicutes	Bacilli	*Sporosarcina*	Chelius and Triplett (2001)
Firmicutes	Bacilli	*Staphylococcus*	Bodhankar et al. (2017), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b, 2017), Mashiane et al. (2017), McInroy and Kloepper (1995b), Orole and Adejumo (2011), Prischl et al. (2012)

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
Firmicutes	Bacilli	*Streptococcus*	Mashiane et al. (2017)
Firmicutes	Clostridia	*Alkaliphilus*	Mashiane et al. (2017)
Firmicutes	Clostridia	*Clostridium*	Johnston-Monje and Raizada (2011)
Firmicutes	Clostridia	*Desulfotomaculum*	Mashiane et al. (2017)
Firmicutes	Clostridia	*Finegoldia*	Liu et al. (2012a)
Firmicutes	Clostridia	*Ruminococcus*	Liu et al. (2012a)
Proteobacteria	Acidithiobacillia	*Acidithiobacillus*	Liu et al. (2012a)
Proteobacteria	Alphaproteobacteria	*Acidiphilium*	Matsumura et al. (2015)
Proteobacteria	Alphaproteobacteria	*Ancylobacter*	Liu et al. (2012a), da Silva et al. (2014) Liu et al. (2012a)
Proteobacteria	Alphaproteobacteria	*Asticcacaulis*	Chelius and Triplett (2001)
Proteobacteria	Alphaproteobacteria	*Aureimonas*	Mashiane et al. (2017)
Proteobacteria	Alphaproteobacteria	*Azospirillum*	Arruda et al. (2013), Estrada et al. (2002), Johnston-Monje and Raizada (2011), Matsumura et al. (2015), Montañez et al. (2008), Orole and Adejumo (2011), Riggs et al. (2001), Roesch et al. (2008)
Proteobacteria	Alphaproteobacteria	*Blastomonas*	Matsumura et al. (2015)
Proteobacteria	Alphaproteobacteria	*Bosea*	Chelius and Triplett (2001), Liu et al. (2012a)
Proteobacteria	Alphaproteobacteria	*Bradyrhizobium*	Chelius and Triplett (2001), Johnston-Monje and Raizada (2011), Liu et al. (2012b), Roesch et al. (2008), Zinniel et al. (2002)
Proteobacteria	Alphaproteobacteria	*Brevundimonas*	Liu et al. (2012a, b), Mashiane et al. (2017), Matsumura et al. (2015), Montañez et al. (2008, 2012), Prischl et al. (2012)
Proteobacteria	Alphaproteobacteria	*Caulobacter*	Chelius and Triplett (2001), Prischl et al. (2012)
Proteobacteria	Alphaproteobacteria	*Devosia*	Liu et al. (2012a), Prischl et al. (2012)
Proteobacteria	Alphaproteobacteria	*Dongia*	Chelius and Triplett (2001)
Proteobacteria	Alphaproteobacteria	*Ensifer*	Prischl et al. (2012)
Proteobacteria	Alphaproteobacteria	*Gluconobacter*	Matsumura et al. (2015)
Proteobacteria	Alphaproteobacteria	*Kaistia*	Chelius and Triplett (2001)
Proteobacteria	Alphaproteobacteria	*Labrys*	Chelius and Triplett (2001)
Proteobacteria	Alphaproteobacteria	*Mesorhizobium*	Chelius and Triplett (2001), Johnston-Monje et al. (2014)
Proteobacteria	Alphaproteobacteria	*Methylobacterium*	Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b), Mashiane et al. (2017), Matsumura et al. (2015), McInroy and Kloepper (1995b), Prischl et al. (2012), Roesch et al. (2008)
Proteobacteria	Alphaproteobacteria	*Methylocystis*	Roesch et al. (2008)
Proteobacteria	Alphaproteobacteria	*Methylosinus*	Roesch et al. (2008)
Proteobacteria	Alphaproteobacteria	*Microvirga*	Liu et al. (2012a)

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
Proteobacteria	Alphaproteobacteria	*Novosphingobium*	Matsumura et al. (2015)
Proteobacteria	Alphaproteobacteria	*Oceanibaculum*	(Liu et al. 2012a)
Proteobacteria	Alphaproteobacteria	*Ochrobactrum*	da Silva et al. (2014), Liu et al. (2012b, 2017)
Proteobacteria	Alphaproteobacteria	*Paracraurococcus*	da Silva et al. (2014)
Proteobacteria	Alphaproteobacteria	*Paracoccus*	Mashiane et al. (2017)
Proteobacteria	Alphaproteobacteria	*Parvibaculum*	da Silva et al. (2014)
Proteobacteria	Alphaproteobacteria	*Pedomicrobium*	da Silva et al. (2014)
Proteobacteria	Alphaproteobacteria	*Phenylobacterium*	Chelius and Triplett (2001)
Proteobacteria	Alphaproteobacteria	*Phreatobacter*	Mashiane et al. (2017)
Proteobacteria	Alphaproteobacteria	*Phyllobacterium*	da Silva et al. (2014), McInroy and Kloepper (1995b), Prischl et al. (2012)
Proteobacteria	Alphaproteobacteria	*Rhizobium/Agrobacterium*	Arruda et al. (2013), Chelius and Triplett (2001), da Silva et al. (2014), Gutiérrez-Zamora and Martínez-Romero (2001), Johnston-Monje et al. (2014), Johnston-Monje and Raizada (2011), Liu et al. (2012a, b), Matsumura et al. (2015), McInroy and Kloepper (1995b), Menéndez et al. (2016), Montañez et al. (2008, 2012), Prischl et al. (2012), Roesch et al. (2008), Rosenblueth and Martinez-Romero (2004), Zinniel et al. (2002)
Proteobacteria	Alphaproteobacteria	*Rhodoblastus*	Roesch et al. (2008)
Proteobacteria	Alphaproteobacteria	*Rhodopseudomonas*	Mashiane et al. (2017)
Proteobacteria	Alphaproteobacteria	*Sandaracinobacter*	Matsumura et al. (2015)
Proteobacteria	Alphaproteobacteria	*Shinella*	Liu et al. (2012a)
Proteobacteria	Alphaproteobacteria	*Sinorhizobium*	Prischl et al. (2012), Roesch et al. (2008), Sandhya et al. (2017)
Proteobacteria	Alphaproteobacteria	*Sphingobium*	Johnston-Monje et al. (2014), Matsumura et al. (2015), Prischl et al. (2012)
Proteobacteria	Alphaproteobacteria	*Sphingomonas*	(Chelius and Triplett 2001), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b, 2017), Mashiane et al. (2017), Matsumura et al. (2015), McInroy and Kloepper (1995b), Prischl et al. (2012), Rijavec et al. (2007)
Proteobacteria	Alphaproteobacteria	*Sphingopyxis*	Liu et al. (2012a)
Proteobacteria	Alphaproteobacteria	*Sphingosinicella*	Chelius and Triplett (2001), Liu et al. (2012a)
Proteobacteria	Alphaproteobacteria	*Xanthobacter*	Matsumura et al. (2015) Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Georgfuchsia*	Liu et al. (2012b)
Proteobacteria	Betaproteobacteria	*Achromobacter*	Arruda et al. (2013), da Silva et al. (2014), Liu et al. (2012b), Pereira et al. (2011)
Proteobacteria	Betaproteobacteria	*Acidovorax*	Liu et al. (2012a, b), Prischl et al. (2012)
Proteobacteria	Betaproteobacteria	*Alcaligenes*	Roesch et al. (2008)

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
Proteobacteria	Betaproteobacteria	*Aquabacterium*	Mashiane et al. (2017)
Proteobacteria	Betaproteobacteria	*Azoarcus*	Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Azohydromonas*	Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Azonexus*	Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Bordetella*	da Silva et al. (2014)
Proteobacteria	Betaproteobacteria	*Burkholderia*	Arruda et al. (2013), Chelius and Triplett (2001), da Silva et al. (2014), Estrada et al. (2002), Ikeda et al. (2013), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b), Mashiane et al. (2017), Matsumura et al. (2015), McInroy and Kloepper (1995b), Montañez et al. (2012), Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Caenimonas*	da Silva et al. (2014)
Proteobacteria	Betaproteobacteria	*Comamonas*	Chelius and Triplett (2001), da Silva et al. (2014)
Proteobacteria	Betaproteobacteria	*Cupriavidus*	Johnston-Monje et al. (2014), Liu et al. (2012a), Johnston-Monje et al. (2014)
Proteobacteria	Betaproteobacteria	*Curvibacter*	Liu et al. (2012a, b)
Proteobacteria	Betaproteobacteria	*Dechloromonas*	Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Delftia*	Johnston-Monje et al. (2014), Liu et al. (2017), Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Derxia*	Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Duganella*	da Silva et al. (2014)
Proteobacteria	Betaproteobacteria	*Herbaspirillum*	Arruda et al. (2013), Chelius and Triplett (2001), da Silva et al. (2014), Johnston-Monje et al. (2014), Montañez et al. (2008, 2012), Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Hydrogenophaga*	McInroy and Kloepper (1995b)
Proteobacteria	Betaproteobacteria	*Ideonella*	Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Inhella*	da Silva et al. (2014)
Proteobacteria	Betaproteobacteria	*Janthinobacterium*	Johnston-Monje et al. (2014)
Proteobacteria	Betaproteobacteria	*Limnobacter*	da Silva et al. (2014), Liu et al. (2012b, 2017)
Proteobacteria	Betaproteobacteria	*Massilia*	Chelius and Triplett (2001), da Silva et al. (2014), Liu et al. (2012a, b), Mashiane et al. (2017)
Proteobacteria	Betaproteobacteria	*Methylibium*	Liu et al. (2012b)
Proteobacteria	Betaproteobacteria	*Methylophilus*	Liu et al. (2012b)
Proteobacteria	Betaproteobacteria	*Methylotenera*	Mashiane et al. (2017)
Proteobacteria	Betaproteobacteria	*Mitsuaria*	Chelius and Triplett (2001)
Proteobacteria	Betaproteobacteria	*Naxibacter*	da Silva et al. (2014)

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
Proteobacteria	Betaproteobacteria	*Pandoraea*	Liu et al. (2012b), Johnston-Monje et al. (2014)
Proteobacteria	Betaproteobacteria	*Pelomonas*	Liu et al. (2012a, b, 2017), Prischl et al. (2012), Roesch et al. (2008)
Proteobacteria	Betaproteobacteria	*Polaromonas*	Chelius and Triplett (2001)
Proteobacteria	Betaproteobacteria	*Pusillimonas*	da Silva et al. (2014)
Proteobacteria	Betaproteobacteria	*Ralstonia*	Chelius and Triplett (2001), Liu et al. (2017)
Proteobacteria	Betaproteobacteria	*Roseateles*	Liu et al. (2012a)
Proteobacteria	Betaproteobacteria	*Silvimonas*	da Silva et al. (2014)
Proteobacteria	Betaproteobacteria	*Tepidimonas*	Liu et al. (2012a)
Proteobacteria	Betaproteobacteria	*Thiobacillus*	Liu et al. (2012a)
Proteobacteria	Betaproteobacteria	*Thiobacter*	Liu et al. (2012b)
Proteobacteria	Betaproteobacteria	*Undibacterium*	Liu et al. (2012b)
Proteobacteria	Betaproteobacteria	*Variovorax*	Chelius and Triplett (2001), Liu et al. (2012a), Mashiane et al. (2017), McInroy and Kloepper (1995b), Prischl et al. (2012)
Proteobacteria	Deltaproteobacteria	*Bdellovibrio*	Liu et al. (2012a)
Proteobacteria	Deltaproteobacteria	*Byssovorax*	Mashiane et al. (2017)
Proteobacteria	Gammaproteobacteria	*Acinetobacter*	Arruda et al. (2013), da Silva et al. (2014), Liu et al. (2012a, b), Mashiane et al. 2017), Matsumura et al. (2015), Liu et al. (2017), Matsumura et al. (2015), Prischl et al. (2012), Sandhya et al. (2017)
Proteobacteria	Gammaproteobacteria	*Aspromonas*	da Silva et al. (2014)
Proteobacteria	Gammaproteobacteria	*Azotobacter*	Roesch et al. (2008)
Proteobacteria	Gammaproteobacteria	*Cedecea*	Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Cellvibrio*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Citrobacter*	Arruda et al. (2013), Johnston-Monje and Raizada (2011), McInroy and Kloepper (1995b), Orole and Adejumo (2011)
Proteobacteria	Gammaproteobacteria	*Cronobacter*	Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Dokdonella*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Enhydrobacter*	Liu et al. (2012a, b)
Proteobacteria	Gammaproteobacteria	*Enterobacter*	Arruda et al. (2013), Chelius and Triplett (2001), da Silva et al. (2014), Estrada et al. (2002), Fisher et al. (1992), Gao et al. (2004), Liu et al. (2012a, b, 2017), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Mashiane et al. (2017), McInroy and Kloepper (1995b), Menéndez et al. (2016), Montañez et al. (2012) Orole and Adejumo (2011) Pereira et al. (2011), Riggs et al. (2001), Sandhya et al. (2017), Seghers et al. (2004), Szilagyi-Zecchin et al. (2014), Zinniel et al. (2002)

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
Proteobacteria	Gammaproteobacteria	*Erwinia*	(Liu et al. 2012a), Pereira et al. (2011), Zinniel et al. (2002)
Proteobacteria	Gammaproteobacteria	*Escherichia*	Johnston-Monje and Raizada (2011), Liu et al. (2012a, b), Matsumura et al. (2015), McInroy and Kloepper (1995b), Zinniel et al. (2002)
Proteobacteria	Gammaproteobacteria	*Hafnia*	Johnston-Monje and Raizada (2011)
Proteobacteria	Gammaproteobacteria	*Halomonas*	Liu et al. (2012a, 2017)
Proteobacteria	Gammaproteobacteria	*Hydrocarboniphaga*	Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Klebsiella*	Arruda et al. (2013), Chelius et al. (2000), Chelius and Triplett (2000), Estrada et al. (2002), Fisher et al. (1992), Ikeda et al. (2013), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. 2012a, b), Mashiane et al. (2017), McInroy and Kloepper (1995b), Palus et al. (1996), Pereira et al. (2011), Roesch et al. (2008), Zinniel et al. (2002)
Proteobacteria	Gammaproteobacteria	*Kluyvera*	McInroy and Kloepper (1995b)
Proteobacteria	Gammaproteobacteria	*Kosakonia*	Menéndez et al. (2016)
Proteobacteria	Gammaproteobacteria	*Leclercia*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Lelliottia*	Matsumura et al. (2015), Menéndez et al. (2016)
Proteobacteria	Gammaproteobacteria	*Luteibacter*	Chelius and Triplett (2001), Johnston-Monje and Raizada (2011), Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Luteimonas*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Lysobacter*	Johnston-Monje et al. (2014), da Silva et al. (2014), Prischl et al. (2012)
Proteobacteria	Gammaproteobacteria	(*Candidatus*) Moranella	Mashiane et al. (2017)
Proteobacteria	Gammaproteobacteria	*Pantoea*	Arruda et al. (2013), Ikeda et al. (2013), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b, Liu et al. 2017), Mashiane et al. (2017), McInroy and Kloepper (1995b), Montañez et al. (2008), (2012), Menéndez et al. (2016), Prischl et al. (2012), Riggs et al. (2001), Rijavec et al. (2007), Sheibani-Tezerji et al. (2015)
Proteobacteria	Gammaproteobacteria	*Perlucidibaca*	Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Providencia*	Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Pseudomonas*	Arruda et al. (2013) Chelius and Triplett (2001), da Silva et al. (2014), Fisher et al. (1992), Gao et al. (2004), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012a, b, 2017), McInroy and Kloepper (1995b, Montañez et al. (2016, 2008), (2012), Orole and Adejumo (2011), Pereira et al.

(continued)

Table 14.1 (continued)

Phylum	Class	Genus[a]	References
			(2011), Prischl et al. (2012), Rijavec et al. (2007), Roesch et al. (2008), Sandhya et al. (2017), Seghers et al. (2004), Zinniel et al. (2002)
Proteobacteria	Gammaproteobacteria	*Pseudoxanthomonas*	Chelius and Triplett (2001), Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Psychrobacter*	Liu et al. (2012a, b)
Proteobacteria	Gammaproteobacteria	*Rahnella*	Mashiane et al. (2017), Montañez et al. (2008, 2012), Menéndez et al. (2016), Seghers et al. (2004)
Proteobacteria	Gammaproteobacteria	*Raoultella*	Roesch et al. (2008)
Proteobacteria	Gammaproteobacteria	*Rheinheimera*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Rhodanobacter*	da Silva et al. (2014)
Proteobacteria	Gammaproteobacteria	*Salinivibrio*	Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Serratia*	Arruda et al. (2013), Gao et al. (2004), Liu et al. (2012a, b), McInroy and Kloepper (1995b)
Proteobacteria	Gammaproteobacteria	*Shigella*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Stenotrophomonas*	Arruda et al. (2013), Chelius and Triplett (2001), da Silva et al. (2014), Johnston-Monje et al. (2014), Johnston-Monje and Raizada (2011), Liu et al. (2012a), McInroy and Kloepper (1995b), Pereira et al. (2011)
Proteobacteria	Gammaproteobacteria	*Steroidobacter*	Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Tatumella*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Thermomonas*	Liu et al. (2012a)
Proteobacteria	Gammaproteobacteria	*Vibrio*	Fisher et al. (1992), Liu et al. (2012b)
Proteobacteria	Gammaproteobacteria	*Xanthomonas*	Gao et al. (2004), Johnston-Monje and Raizada (2011), Johnston-Monje et al. (2014), Liu et al. (2012b), McInroy and Kloepper (1995b), Prischl et al. (2012), Zinniel et al. (2002)

[a]This table is limited to organisms which were actually isolated from maize, as opposed to those that have simply been shown capable of growing within maize when artificially inoculated. Genera from (Chelius and Triplett 2001) were identified by BLASTing the NCBI accession numbers against a local copy of the NCBI nucleotide database (downloaded May 15, 2017) and manually identifying a consensus among the top ten BLAST hits

be possible that maize tissues can enter a receptive state to encourage endophyte invasion, analogous to how legume roots are primed to attract rhizobia and initiate nodulation. If such a state exists, however, it would have to be very subtle to have avoided detection until now.

Since many maize endophytes are horizontally transmitted, it may not be surprising that they are also generalists, capable of infecting several plant taxa. In fact, we are unaware of any endophytes found exclusively in maize. Depending on the endophyte, they may also be found in other grasses (Bacon and Hinton 2002), legumes (Bacon and Hinton 2002), cucurbits (Zinniel et al. 2002), euphorbs (Mohanty et al. 2017), and even trees (Knoth et al. 2013). Little is known about the population structures of these generalist endophytes, and understanding them could have important consequences for agriculture. For example, we might expect the

Table 14.2 Known maize endophytes (fungi)

Phylum	Class	Genus[a]	References
Ascomycota	–	*Verticillium*	Fisher et al. (1992)
Ascomycota	Ascomycetes	*Cochliobolus*	Bokati et al. (2016), Pan et al. (2008)
Ascomycota	Dothideomycetes	*Alternaria*	Bokati et al. (2016), Brookes (2017), Fisher et al. (1992), Orole and Adejumo (2009, 2011), Pan et al. (2008)
Ascomycota	Dothideomycetes	*Aureobasidium*	Fisher et al. (1992), Pan et al. (2008)
Ascomycota	Dothideomycetes	*Cladosporium*	Brookes (2017), Fisher et al. (1992), Pan et al. (2008)
Ascomycota	Dothideomycetes	*Darksidea*	Bokati et al. (2016)
Ascomycota	Dothideomycetes	*Exserohilum*	Bokati et al. (2016)
Ascomycota	Dothideomycetes	*Leptosphaeria/Phaeosphaeria*	Pan et al. (2008)
Ascomycota	Dothideomycetes	*Lewia*	Pan et al. (2008)
Ascomycota	Dothideomycetes	*Paraconiothyrium*	Bokati et al. (2016)
Ascomycota	Dothideomycetes	*Phoma*	Bokati et al. (2016), Orole and Adejumo (2009), Pan et al. 2008)
Ascomycota	Dothideomycetes	*Epicoccum*	Brookes (2017), Fisher et al. (1992), Pan et al. (2008)
Ascomycota	Euascomycetes	*Curvularia*	Pan et al. (2008)
Ascomycota	Eurotiomycetes	*Aspergillus*	Amin (2013), Orole and Adejumo (2011)
Ascomycota	Eurotiomycetes	*Penicillium*	Amin (2013), Brookes (2017), Fisher et al. (1992)
Ascomycota	Eurotiomycetes	*Talaromyces*	Peterson and Jurjević (2017)
Ascomycota	Saccharomycetes	*Saccharomyces*	Orole and Adejumo (2011)
Ascomycota	Saccharomycetes	*Williopsis*	Nassar et al. (2005)
Ascomycota	Sordariomycetes	*Acremonium*	Amin (2013), Fisher et al. (1992), Gams (1971), Orole and Adejumo (2009, 2011), Pan et al. (2008)
Ascomycota	Sordariomycetes	*Beauveria*	Orole and Adejumo (2009b)
Ascomycota	Sordariomycetes	*Botryodiplodia*	Amin (2013)
Ascomycota	Sordariomycetes	*Chaetomium*	Bokati et al. (2016), Fisher et al. (1992)
Ascomycota	Sordariomycetes	*Colletotrichum*	Orole and Adejumo (2011), Pan et al. (2008)
Ascomycota	Sordariomycetes	*Fusarium*	Amin (2013), Bacon and Hinton (1996), Bokati et al. (2016), Brookes (2017), Fisher et al. (1992), Leslie et al. (1990), Orole and Adejumo (2011), Pan et al. (2008)
Ascomycota	Sordariomycetes	*Gibberella*	Bokati et al. (2016)
Ascomycota	Sordariomycetes	*Microdochium*	Fisher et al. (1992)
Ascomycota	Sordariomycetes	*Podospora*	Bokati et al. (2016)

(continued)

Table 14.2 (continued)

Phylum	Class	Genus[a]	References
Ascomycota	Sordariomycetes	*Trichoderma*	Amin (2013), Brookes (2017), Fisher et al. (1992), Orole and Adejumo (2009, (2011)
Basidiomycota	Agaricomycetes	*Ceratobasidium*	Bokati et al. (2016)
Basidiomycota	Microbotryomycetes	*Sporobolomyces*	Pan et al. (2008)
Basidiomycota	Pucciniomycetes	*Puccinia*	Pan et al. (2008)
Basidiomycota	Tremellomycetes	*Bullera/Bulleromyces*	Pan et al. (2008)
Basidiomycota	Tremellomycetes	*Cryptococcus*	Pan et al. (2008)
Basidiomycota	Tremellomycetes	*Filobasidium*	Pan et al. (2008)
Basidiomycota	Urediniomycetes	*Rhodotorula*	Nassar et al. (2005), Pan et al. (2008)
Basidiomycota	Ustilaginomycetes	*Ustilago*	Fisher et al. (1992), Pan et al. (2008)
Zygomycota	Mucormycotina	*Mucor*	Brookes (2017)

[a]This table is limited to organisms which were actually isolated from maize, as opposed to those that have simply been shown capable of growing within maize when artificially inoculated. Several genera with known pathogens are included (e.g., *Fusarium*, *Ustilago*, *Verticillium*) based on the original authors' classification as endophytes. Some of these may have been truly endophytic (i.e., nonpathogenic), while others may have been latent or only in the early stages of infection

Table 14.3 Known maize endophytes (archaea)

Phylum	References
Euryarchaeota	(Chelius and Triplett 2001)
Thaumarchaeota[a]	(Chelius and Triplett 2001)

[a]The original publication lists this as "Marine Group I" Crenarchaeota. Localization within the Thaumarchaeota was confirmed with leBIBI^{QBPP} (Flandrois et al. 2015)

monocultures of modern agriculture to favor endophytes evolving specialization and parasitism, while more diverse natural communities could favor less harmful, generalist populations of endophytes (e.g., Leggett et al. 2013).

14.2.3 Localization

Endophytes have been identified in all parts of the maize plant, including roots (Fisher et al. 1992; Bacon and Hinton 1996, 2002; Chelius and Triplett 2000; Roesch et al. 2008; Ikeda et al. 2013; Naveed et al. 2014), stalks (Fisher et al. 1992; Bacon and Hinton 1996, 2002; Chelius and Triplett 2000; Roesch et al. 2008), leaves (Fisher et al. 1992; Bacon and Hinton 1996, 2002; Naveed et al. 2014), and seeds (Fisher et al. 1992; McInroy and Kloepper 1995a; Johnston-Monje and Raizada 2011; Liu et al. 2012b, 2015b; Gond et al. 2015). There are no reports of endophytes in maize pollen, but given its large size and the presence of endophytes in the pollen of other plants (Madmony et al. 2005; Hodgson et al. 2014), it seems likely that they exist.

Most maize endophytes do not live inside plant cells themselves but rather in the intercellular spaces (apoplast) (Chelius and Triplett 2000; Bacon and Hinton 2002; Nassar et al. 2005; Liu et al. 2006) or vasculature (Roncato-Maccari et al. 2003; Nassar et al. 2005; Liu et al. 2006; Johnston-Monje and Raizada 2011). Some endophytes are confined to specific tissues (frequently roots; Harman et al. 2004; Roesch et al. 2008), while many can grow systemically throughout the plant (Bacon and Hinton 1996; Lamb et al. 1996; Roncato-Maccari

et al. 2003; Liu et al. 2006; Naveed et al. 2014). In the latter case, their abundance often decreases further away from the roots (Fisher et al. 1992; Lamb et al. 1996; Bacon and Hinton 2002; Roesch et al. 2006, 2008; Naveed et al. 2014). Some endophytes apparently grow or move from the plant interior onto the plant surface (Lamb et al. 1996) or into the rhizosphere (Johnston-Monje and Raizada 2011; Johnston-Monje et al. 2016), implying that infected seeds can be a source of local inoculum for other plants.

14.2.4 Abundances Within Plants

It is surprisingly difficult to estimate how many endophytes live inside an individual maize plant. Most estimates are made by grinding tissue and plating dilutions to estimate the number of colony-forming units (CFUs) in the original material. This suffers the obvious bias of only counting what can grow under laboratory conditions, but has the advantage of being easy and relatively robust. Quantitative (Rodriguez Estrada et al. 2011) or semiquantitative (Johnston-Monje and Raizada 2011) PCR can also be used, but usually only gives relative biomass estimates and is limited to evaluating those taxa for which primers are designed. Direct counting, meanwhile, requires tedious microscopic examination and is rarely used. In short, no one has yet developed a truly accurate method to quantify endophytes, so all estimates should be taken as tentative.

Due to these and other methodological limitations, estimates of endophyte loads in maize plants vary over several orders of magnitude. Bacterial endophytes have been counted at 10^3–10^{10} CFUs per gram of fresh maize shoot and 10^4–10^{10} CFUs per gram of root; both of these counts are probably underestimates (McInroy and Kloepper 1991, 1995a; Estrada et al. 2002). Plants grown without soil show much lower bacteria counts ($\sim 10^2$ CFUs per gram fresh weight; McInroy and Kloepper 1995a), suggesting that the environment may play a role in determining endophyte load. However, plants grown in sterile sand can still show similar bacterial *diversity* as soil-grown plants (Johnston-Monje et al. 2016). Fungal endophyte counts are much harder to acquire than bacterial ones, and few authors have tried to directly estimate fungal biomass in living plants. One example is that of a maize pathogen, *Aspergillus flavus*, which was found in maize kernels at rates of 1.9–7.3 mg hyphae per gram tissue (Williams et al. 2011). This is probably an upper limit for endophyte colonization, since one would expect a pathogen to multiply to a much higher level than a nonpathogenic endophyte.

There appear to be no published studies quantifying the number of distinct endophytes in a single maize plant, though unpublished data estimates 30–40 unique fungal species and 150–200 bacterial species per plant (Barry Goldman, personal communication). Most authors tend to aggregate results across plants, but their numbers are roughly in line with this (de Araujo et al. 2000; Zinniel et al. 2002; Seghers et al. 2004; Johnston-Monje and Raizada 2011; da Silva et al. 2014; Mousa et al. 2015). Interestingly, Ali et al. (2017) found that highly productive areas within a maize field consistently have both more endophytic species (especially *Pseudomonas* spp.) and more uneven species distributions than low-performing regions, though these may just be indications of local soil conditions.

When taken together, these reports imply a much larger diversity of organisms capable of living in maize than are actually in any individual plant. To fully understand how endophytes colonize maize, we will need to better define ecological niches within plants, understand why certain species occupy these niches given the pool available from the environment, and determine how interactions among endophytes affect their community structure.

14.3 The Effects of Endophytes on Maize Plants

Many endophytes are not just passive inhabitants of maize, but may instead have significant beneficial or harmful impacts on their plant host. The nature of these impacts varies widely by microbial species and even by strain within a species.

14.3.1 Growth Effects

One of the most commonly reported effects of endophytes on maize is that of growth promotion, usually measured by increased biomass (Gutierrez-Zamora and Martinez-Romero 2001; Riggs et al. 2001; Bacon and Hinton 2002; Rosenblueth and Martinez-Romero 2004; Nassar et al. 2005; Hungria et al. 2010; Canellas et al. 2012; Montañez et al. 2012; Baldotto et al. 2012; Arruda et al. 2013; Knoth et al. 2013; Young et al. 2013; Johnston-Monje et al. 2014; Akhtar et al. 2015; Mohanty et al. 2017). The prevalence of growth-promoting phenotypes reported in the literature probably reflects the interests of maize researchers more than it does any general property of endophyte communities. There are presumably many endophytes that do not boost growth or even retard it, but these are only rarely reported. For example, only 2 out of 91 maize endophytes boosted growth of another plant (potato shoots), while more than 50 stunted it (Johnston-Monje and Raizada 2011), implying that growth-promoting organisms may actually be the minority.

Aside from general growth, endophytes have also been shown to alter specific morphological aspects of maize plants, especially in roots. The bacterium *Herbaspirillum seropedicae,* for example, increases the number of lateral roots in infected seedlings 7 days after infection (do Amaral et al. 2014; Ferrari et al. 2014). *Bacillus mojavensis* RCC 101 and *Burkholderia* sp. CC-A174 increase root length (Bacon and Hinton 2002; Young et al. 2013), and *Trichoderma* T22 increases both root length and root hair area (Harman et al. 2004).

14.3.2 Nutrient Acquisition

Several endophytes help supply maize with essential nutrients. The best-studied examples are the diazotrophs capable of fixing atmospheric nitrogen to a usable form (Riggs et al. 2001; Estrada et al. 2002; Montañez et al. 2008; Hungria et al. 2010; Matsumura et al. 2015). Unlike the rhizobia that fix nitrogen in legume nodules, maize-associated diazotrophs live throughout the plant as typical endophytes (Santi et al. 2013). Not surprisingly, they tend to have the largest impact on their host when soil nitrogen is scarce; well-fertilized maize shows no effect or can sometimes even show a decrease in yield when colonized by diazotrophs (Matsumura et al. 2015). Although a maize plant can harbor several different diazotrophs, only a subset of them appear to actually fix nitrogen in plants (Roncato-Maccari et al. 2003).

Various bacterial endophytes have been shown to improve plants' ability to acquire nitrogen, phosphorus, and potassium from the soil, including strains of *Burkholderia* (Young et al. 2013), *Herbaspirillum* (Baldotto et al. 2012), and *Azospirillum* (Hungria et al. 2010). *Azospirillum* was also shown to increase the uptake of micronutrients, including boron, copper, magnesium, manganese, sulfur, and zinc (Hungria et al. 2010). Since the endophytes are not in contact with the soil themselves, the increased nutrient status is probably an indirect result of other effects, such as changes in root size or morphology.

14.3.3 Diseases

As usually defined, endophytes do not cause disease themselves. They can, however, either prevent or facilitate infection by other disease-causing organisms (Arnold et al. 2003). The overall prevalence of maize endophytes that are disease-inhibiting versus disease-facilitating is not known, although a recent study in poplar found roughly equal numbers of both (Busby et al. 2016).

Nonetheless, much research has profitably focused on identifying endophytes that antagonize pathogens for their potential use in biocontrol (Hawkes and Connor 2017). Endophytic bacteria have been isolated that antagonize *Fusarium* pathogens (Bacon and Hinton 2002; Mousa et al. 2015), including reducing lesion size (Bacon and Hinton 2011) and different rot symptoms (Orole and Adejumo 2009). In addition, nonpathogenic strains of *Fusarium* can be antagonistic toward smut (*Ustilago maydis*) (Lee et al. 2009), and a strain of *Bacillus subtilis* can inhibit southern leaf blight (Ding et al. 2017). Several species have been isolated with broad antifungal or antipathogen properties (de Araujo et al. 2000; Wicklow et al. 2005; Rijavec et al. 2007; Wicklow and Poling 2009; Chulze et al. 2015; Gond et al. 2015; Liu et al. 2015b; Mousa et al. 2015; Sheibani-Tezerji et al. 2015; Shehata et al. 2016, 2017; Potshangbam et al. 2017; Shehata and Raizada 2017), so that a single endophyte could conceivably protect against many diseases, even including human food-borne illnesses (Shehata et al. 2017). Other protective effects are indirect, for example the induction of systemic plant resistance by *Trichoderma* (Harman et al. 2004). In this case, even though the fungus is confined to the roots, it induces resistance to a pathogen (anthracnose) in the leaves. Exploiting these sorts of interactions is currently an active area of "phytobiome" research for crops (Hawkes and Connor 2017) and one that bears further investigating. For example, theory suggests that competition between endophytes and pathogens could select for endophytes with greater virulence toward the host, though the exact outcome depends on local conditions (Nelson and May 2017). Better understanding these interactions under natural conditions should help us understand (and influence) the long-term consequences of rolling out endophyte-based biocontrol.

Situations where endophytes facilitate infection in maize by other organisms are much less well studied. One of the few examples is that of *Fusarium verticillioides*, which is usually considered a pathogen but may exist more broadly as an endophyte (Kuldau and Yates 2000; Bacon et al. 2008; Pan and May 2009). *F. verticillioides* can break down the plant defensive compound BOA (Glenn et al. 2001, 2002; Saunders et al. 2010), and through this and possibly other mechanisms, it facilitates the infection and growth of both commensal endophytic and potential disease-causing fungi (Saunders and Kohn 2008). A study of poplar endophytes, meanwhile, showed that endophytic fungi that promoted disease were common and were sometimes closely related to ones that provided protection (Busby et al. 2016).

Sometimes context can change whether an endophyte inhibits or enhances a pathogen. For example, Lee et al. (2009) found that simultaneous co-inoculation of *Fusarium verticillioides* with *Ustilago maydis* reduced the severity of disease caused by *U. maydis*. However, the effect disappeared if there was a 2-day gap between infections, regardless of which organism was introduced first. In contrast, Sobowale et al. (2007) identified several species of *Trichoderma* capable of outcompeting pathogenic *F. verticillioides* even when the pathogen was given a 24-h head start. These results suggest that microbe–microbe interactions in the maize endosphere can be very complex and depend not just on the microbes involved but also on the order they arrive and, presumably, other aspects of the local microenvironment.

14.3.4 Insect Pests

Only a few maize endophytes have been studied with regard to their effect on insect pests. The best-studied of these is the fungus *Beauveria bassiana*, which has been shown by several authors to be antagonistic toward maize stem borers (Lewis and Cossentine 1986; Cherry et al. 1999, 2004). Although the *Beauveria* genus contains known insect pathogens, stem borer suppression in maize is thought to rely on fungal metabolites produced in the plant instead of infection of the insect (Wagner and Lewis 2000; Cherry et al. 2004). Similarly, root-associated *Trichoderma atroviride* is thought to reduce damage from fall armyworm by producing

volatile organic compounds that deter the insects and/or induce plant defenses (Contreras-Cornejo et al. 2017). Finally, maize roots colonized by *Azospirillum brasilense* showed reduced feeding and growth of corn rootworm (Santos et al. 2014), although it was unclear whether the bacteria had colonized the root interior or just the surface.

14.3.5 Abiotic Stress Tolerance

Surprisingly few maize endophytes have been evaluated for their effect on plant abiotic stress tolerance. Endophyte-mediated drought resistance (Casanovas et al. 2002; Cohen et al. 2009; Naveed et al. 2014; Sandhya et al. 2017) is thought to be at least partly due to the endophyte producing plant hormones such as abscisic acid and gibberellins (Cohen et al. 2009). Production of compatible solutes also likely plays a role, and many non-endophytic rhizobacteria have been shown to influence the accumulation of different solutes (reviewed in Vurukonda et al. 2016). Endophytes have also been shown to boost resistance to salinity stress (Akhtar et al. 2015) and to heavy metals such as cadmium, zinc, and lead (Li et al. 2011; Wang et al. 2016; Ban et al. 2017; He et al. 2017). Heavy-metal resistance mechanisms include boosting antioxidant activity, conversion of metals to nontoxic forms, sequestering metals in cell walls and in roots, and altering phytohormone levels to overcome toxicity.

14.4 Assembly of the Maize Endophyte Community

Maize–endophyte interactions are ultimately governed by a combination of plant, endophyte, and environment, and the connections among these factors are undoubtedly complex. Achieving a systems-level understanding of the dynamics of plant–endophyte interactions will be an important part of both understanding maize endophytes and harnessing them for global agriculture.

14.4.1 Endophyte Acquisition and Transmission

Although some maize endophytes are transmitted via seed (McInroy and Kloepper 1995a; Johnston-Monje and Raizada 2011; Liu et al. 2012b, 2015b; Johnston-Monje et al. 2014), many appear to be acquired from the local environment, especially the soil (McInroy and Kloepper 1995a; Bokati et al. 2016). This means that much of the maize endophyte community within each plant is built anew every generation.

A fruitful avenue of investigation will be to determine the mechanisms governing endophyte community assembly. These are largely unknown but may vary by tissue even within a single plant. For example, Johnston-Monje and Raizada (2011) found that the endophytes of maize seeds were phylogenetically clustered by host (teosintes, ancient landraces, modern landraces, and improved varieties) even after growing the genotypes in a common garden. However, there was significant turnover evidenced by changes in 54–87% of the bacterial taxa over a single generation. In contrast, stem endophytes showed no phylogenetic clustering, suggesting that a broad array of taxa is redundant in function. A follow-up study using different soils found that the different plant genotypes assembled different communities across many substrates, including sterile sand (Johnston-Monje et al. 2014). In this case, the host genotype had a larger impact on endophyte community than did the inoculum source (soil), while tissue (root versus shoot) had an even larger impact than genotype.

It is tempting to argue that maize and its ancestor teosinte evolved to work with specific endophytes, but that the endophytes did not move with maize out of its native range in central Mexico. Mousa et al. (2015) purport evidence for this hypothesis, finding that the strongest antagonists of pathogens all came from wild teosinte. However, spatial ecological and evolutionary processes may be confounded in such studies. Meanwhile, Bokati et al. (2016) found no significant differences between maize and teosinte when grown in either clay or desert soils,

implying that domestication has not significantly altered the assembly of endophytic communities in maize. Since maize endophyte communities are similar across very different substrates (Johnston-Monje et al. 2014), it seems likely that maize can assemble similar endophyte communities from many different starting points. Even if there are specific teosinte endophytes that would be beneficial for maize, it seems unlikely that general maize endophyte communities in Mexico are inherently superior to those outside it.

14.4.2 Endophyte Influences on Their Plant Host

Endophytes live in intimate association with their maize host. Although one might expect endophytes to evolve to avoid immune detection entirely, it appears that they actually trigger immunity at low levels (Van Wees et al. 2008; Reinhold-Hurek and Hurek 2011). This low-level immune response may constitute one mechanism of host control on endophytes and may have the collateral effect of "priming" plants so that they better fend off actual pathogens.

Endophytes are also privy to internal plant signals such as hormones, sugar levels, and metabolites. Many endophytes have been found to manipulate these signals, with synthesis of growth hormones particularly common. Endophytes can synthesize auxins (especially indole-3-acetic acid) (Nassar et al. 2005; Montañez et al. 2012; Szilagyi-Zecchin et al. 2014; Menéndez et al. 2016; Sandhya et al. 2017; Potshangbam et al. 2017), gibberellin (Lucangeli and Bottini 1997; Cohen et al. 2009; Sandhya et al. 2017), abscisic acid (Cohen et al. 2009), cytokines (Sandhya et al. 2017), ACC deaminase (Menéndez et al. 2016; Sandhya et al. 2017), and other growth regulators (Gold et al. 2014). These manipulations sometimes benefit the plant but presumably always benefit the endophyte.

14.4.3 Plant Influences on Endophytes

The genotype of maize plants is known to modulate endophyte interactions, although most reports are limited to incidental statements of endophytes acting differently in different maize varieties (Riggs et al. 2001; Montañez et al. 2008, 2012; Pan et al. 2008; Prischl et al. 2012; Naveed et al. 2014; Sheibani-Tezerji et al. 2015; Brusamarello-Santos et al. 2017). Johnston-Monje et al. (2014) did find that the effect of maize genetics was greater than soil type on endophytic composition, though the effect of tissue type within maize plant was bigger still. A similar result was found by Mashiane et al. (2017), who found much larger differences in maize endophytic communities due to the growth stage of the plant than to plant genotypes. Because many maize endophytes have broad host ranges (see Sect. 14.2.2; Bacon and Hinton 2002; Knoth et al. 2013), the effect of individual plant genotypes might tune interactions instead of completely allowing or disallowing them.

One way that specific maize genes have been shown to influence their interactions with endophytes is through the production of benzoxazinoid defense compounds. Maize plants that produce these compounds have significantly different fungal endophytes than those that do not, including boosting the colonization of *Fusarium* in ways that may be harmful to crop production (Saunders and Kohn 2009).

14.4.4 Environmental Influences on Endophytic Interactions

Although some work has been done to determine how management practices affect maize endophytes (see Sect. 14.6), little information on the effects of natural environmental variables on

these communities is available. Established endophytes are partly buffered against the outside environment, but temperature, water status, nutrient availability, and other environmental conditions presumably still affect them. Soil type should impact endophytes simply by being a source of inoculum; yet while some studies show a significant impact of soil (Chen et al. 2017), others indicate it has a minimal effect (Johnston-Monje et al. 2014, 2016). Work in other plant species has found effects on endophytes due to water (Zimmerman and Vitousek 2012; Giauque and Hawkes 2013), temperature (Zimmerman and Vitousek 2012), and spatial (David et al. 2016) or geographic location (U'Ren et al. 2012). Such results imply that many environmental factors will affect maize endophyte community diversity and that understanding these factors will likely be important for both basic and applied maize research.

14.4.5 Range of Endophyte Lifestyles

Although endophytes by our definition do not cause apparent disease symptoms, that does not mean they are all beneficial to their host plant. A growing number of studies looking at the effect of endophytes on plant health have shown that different endophytes can have dramatically different impacts on plant health (e.g., Wicklow and Poling 2009; Sheibani-Tezerji et al. 2015). These results suggest that endophytes, as a class, can occupy the entire range of lifestyles from mutualist to commensalist to parasite. Different lifestyles can occur even among strains of the same species. For example, the species *Fusarium verticillioides* includes both endophytes and pathogens (Kuldau and Yates 2000), and it can protect the plant against the smut pathogen *Ustilago maydis* (Lee et al. 2009) but also facilitate infection by other disease-causing organisms (Saunders and Kohn 2009). On a larger phylogenetic scale, the results of Arnold et al. (2009) demonstrate that transitions in trophic modes (symbiosis, pathogenesis, saprobic) may occur frequently within the Ascomycetes, the most common phylum of endophytic fungi, and that transitions between pathogenic and endophytic modes occur equally often in both directions.

14.5 Genomic Contributions of Maize Endophytes

The endophyte lifestyle puts interesting ecological and evolutionary processes into play. Since many maize endophytes are acquired from the local environment (McInroy and Kloepper 1995a; Bokati et al. 2016), they face pressure to adapt to both the external environment and the maize plant interior. These competing selection pressures will undoubtedly leave their mark on endophyte genomes. For example, the genomic structure of rhizobia symbiont populations is strongly affected by adaptation to the soil habitat while still acquiring genes for symbiosis that are specific to host interactions (Hollowell et al. 2016). For our purposes, we will focus on endophyte genomics mostly as they relate to interactions with maize and contributions to the maize extended phenotype. For a more general overview of endophyte genomics, see Frank (2011) and Brader et al. (2017).

14.5.1 Maize Endophyte Genomics

Before getting into specifics, one must first appreciate that the genomic repertoire of endophytes vastly outnumbers that of their host plant. Based on the genomes currently in NCBI, bacteria and fungi contain an average of 3,800 and 9,600 genes, respectively. If we assume 30–40 unique fungi and 150–200 unique bacteria per plant (see Sect. 14.2.4), it implies that the maize endosphere contains 858,000–1.14 million genes, over twenty times that of the maize plant itself. Even allowing for similar genes in many of these organisms, that implies a huge functional diversity available to the maize plant via its endophytes. This massive microbial genetic diversity is analogous to that of the human gut microbiome, which is estimated to harbor over 9 million unique genes, roughly 400 times that of its human host (Yang et al. 2009).

Only a few maize endophytes have had their genomes sequenced (Table 14.4). This is probably an undercount, since maize endophytes may not be identified as such in their genome records, but it drives home the point that there are few genomic resources for these species from which to draw conclusions. There are, however, many genomes for other organisms in genera known to include maize endophytes, although the extent of coverage varies widely by clade (Fig. 14.2). Gene content can vary dramatically within a species (Medini et al. 2005), so these relatives are an imperfect representation at best. Still, what can we learn from these resources about maize endophyte genomics?

First, the functions provided by endophyte genomes are diverse. Some of the more common functions include genes for nitrogen fixation (Riggs et al. 2001; Estrada et al. 2002; Montañez et al. 2008; Fouts et al. 2008; Hungria et al. 2010; Matsumura et al. 2015), hormone synthesis (Nassar et al. 2005; Pedrosa et al. 2011; Weilharter et al. 2011; Montañez et al. 2012; Gold et al. 2014; Sheibani-Tezerji et al. 2015; Menéndez et al. 2016), defensive compound synthesis (Poling et al. 2008; Gold et al. 2014; Gond et al. 2015), and/or defense compound breakdown (Glenn et al. 2001, 2002; Fouts et al. 2008; Saunders et al. 2010; Pedrosa et al. 2011; Weilharter et al. 2011). The latter includes genes for dealing with reactive oxygen species, which endophytes may encounter as part of plant defense responses.

Second, most endophyte genomes contain genes involved in the breakdown of plant physical structures (cellulases, pectin lyases, expansins, etc.), presumably to provide the endophyte with carbon sources once inside the plant (Fouts et al. 2008; Weilharter et al. 2011; Szilagyi-Zecchin et al. 2014; Gold et al. 2014). However, cell wall degrading enzymes may not be required for infection since endophytes can enter through wounds or stomata. *Herbaspirillum seropedicae* SmR1 even lacks these genes entirely and yet can move from soil to maize xylem in less than 24 h (Monteiro et al. 2008; Pedrosa et al. 2011).

Third, maize endophytes frequently have a large repertoire of secretion systems (Fouts et al. 2008; Pedrosa et al. 2011; Sheibani-Tezerji et al. 2015) that may play a role in plant–microbe communication. Unfortunately, there are too few genomes available to draw conclusions about the relative importance of different secretion system types. We do know that Type III systems are missing in *Klebsiella pneumoniae* 342 (Fouts et al. 2008), Type IV are missing in *Herbaspirillum seropedicae* SmR1 (Pedrosa et al. 2011), and both are missing in *Azoarcus* sp. BH72 (Krause et al. 2006). It has been speculated that the loss of specific secretion systems (especially Type III and Type VI) could predispose organisms to adopt endophytic instead of pathogenic lifestyles (Reinhold-Hurek and Hurek 2011; Brader et al. 2017).

14.5.2 Comparisons Between Endophytes and Their Non-endophytic Relatives

Given the diversity of organisms in the endosphere, it seems likely that there are no core "endophyte genes," especially when comparing across bacteria and fungi. However, comparisons between closely related species may identify sets of genes that are related to an endophytic lifestyle, as shown by the few available studies below.

Comparison of a nitrogen-fixing, endophytic strain of *Klebsiella pneumoniae* with a non-endophytic strain that is a potential vertebrate pathogen found several key differences between the two (Fouts et al. 2008). Aside from the obvious difference of nitrogen fixation, the endophyte contained more transcription factors and signal transduction genes, a pair of Type IV secretion systems absent in the non-endophyte and fewer cell surface structures (speculated to be involved in evading the plant immune response).

Sheibani-Tezerji et al. (2015) compared the genomes of three closely related strains of

Table 14.4 Maize endophytes with sequenced genomes

Species[a]	Kingdom	NCBI BioProject	Reference	Original source[b]	Effect on maize
Bacillus mojavensis RRC101	Bacteria (Firmicutes)	PRJNA200690	Gold et al. (2014)	Maize seed (as *Enterobacter cloacae*) (Hinton and Bacon 1995)	Biocontrol of *Fusarium* (Bacon and Hinton 2002, 2011)
Burkholderia phytofirmans PsJN	Bacteria (Betaproteobacteria)	PRJNA17463	Weilharter et al. (2011)	Onion roots (Sessitsch et al. 2005)	Drought tolerance (Naveed et al. 2014)
Klebsiella pneumoniae 342	Bacteria (Enterobacteria)	PRJNA224116	Fouts et al. (2008)	Maize stems (Chelius et al. 2000)	Nitrogen fixation (presumed in maize, confirmed in wheat; Iniguez et al. 2004)
Pantoea ananatis S6, S7, and S8	Bacteria (Enterobacteria)	PRJEB7511 PRJEB7512 PRJEB7513	Sheibani-Tezerji et al. (2015)	Maize seed (Sheibani-Tezerji et al. 2015)	Boost growth (S6), mild pathogen (S7), commensal (S8) (Sheibani-Tezerji et al. 2015)
Rhizobium phaseoli Ch24-10	Bacteria (Alphaproteobacteria)	PRJNA79359	López-Guerrero et al. (2012)	Maize stems (as *Rhizobium etli*) (Rosenblueth and Martinez-Romero 2004)	Not tested; related strain boosts growth (Gutiérrez-Zamora and Martínez-Romero 2001)
Burkholderia gladioli UCD-UG_CHAPALOTE	Bacteria (Betaproteobacteria)	PRJNA260296	Ettinger et al. (2015)	Maize seeds (Johnston-Monje and Raizada 2011)	Not tested; antagonizes fungal pathogens in creeping bentgrass (Shehata 2016)
Herbaspirillum seropedicae Z67	Bacteria (Betaproteobacteria)	PRJNA224116	Not yet published	Rice roots (Baldani et al. 1986)	Improved growth under low nutrients (Canellas et al. 2012)
Herbaspirillum seropedicae SmR1	Bacteria (Betaproteobacteria)	PRJNA47945	Pedrosa et al. (2011)	Sorghum roots (Baldani et al. 1986); strain SmR1 is a spontaneous mutant of strain Z78 (Pedrosa et al. 2011)	Nitrogen fixation and growth promotion (presumed in maize, confirmed in rice; Gyaneshwar et al. 2002)
Serratia sp. *S119*	Bacteria (Gammaproteobacteria)	PRJNA342012	Ludueña et al. (2018)	Peanut nodules (Taurian et al. 2010)	Phosphate acquisition and growth promotion (Ludueña et al. 2017)

[a]Only organisms with confirmed examples of endophytic (not pathogenic) growth are included in this table. This list is probably incomplete, as other maize endophytes may have been sequenced but are not annotated as maize endophytes

[b]All organisms in this table can grow endophytically in maize, although some were originally isolated elsewhere

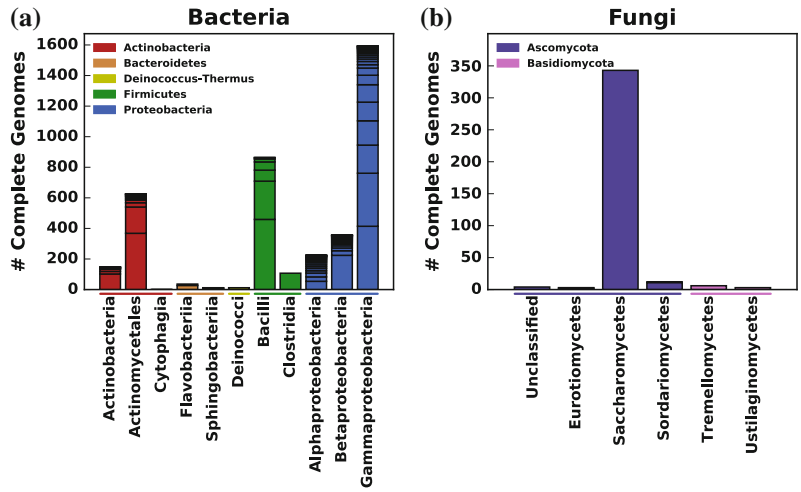

Fig. 14.2 Sequenced genomes of maize endophyte relatives. Each panel shows the number of microbial genomes in NCBI that are in the same genus as known maize endophytes (Tables 14.1 and 14.2) and have at least a chromosome-level assembly at the time of this writing (October 2017). **a** Relatives of bacterial endophytes. **b** Relatives of fungal endophytes. Each graph indicates the phylum by color, the order by stacked bar plots, and genera by individual bars within each stack. Genome counts are highly skewed toward a few clades for both bacteria and fungi, implying that there are still large gaps in our knowledge of maize endophyte genomics

Pantoea ananatis with different lifestyles (pathogenic, commensal, mutualistic). All three genomes were >99% similar to a reference strain (AJI3355) at the nucleotide level and shared 88.9% of protein clusters among all three strains. Protein differences among strains were concentrated in Type VI secretion system genes, effector proteins (bacterial proteins that manipulate plant metabolism; Toruño, Stergiopoulos and Coaker 2016), and parasitic elements (transposases, integrases, and phage genes), and the authors speculate that the latter may relate to the ability to acquire new genetic diversity.

The genome of *Azoarcus* sp. BH72, a general grass endophyte originally isolated from Kallar grass (*Leptochloa fusca*), showed significant differences from *Azoarcus* sp. EbN1, a related soil organism (Krause et al. 2006). These differences include several gene clusters for cell surface components (speculated to be involved in plant–microbe interactions), reduced carbon source utilization, and fewer genes involved in pathogenesis, including a complete lack of Type III and Type IV secretion systems. In contrast, *Herbaspirillum seropedicae* SmR1 contained a multitude of secretion systems, including both Type III system and Type IV pili, though it lacked enzymes to degrade cell walls (Pedrosa et al. 2011).

Comparing the genomes of several endophytes from different plant species reveals several intriguing patterns but no ironclad rules (Mitter et al. 2013). For example, while most endophytes make flagella, some (e.g., *Klebsiella pneumoniae* 342) have lost them entirely. Others (*Azospirillum* sp. B510 and *Gluconacetobacter diazotrophicus* PA15) lack any genes for cell adhesion that are abundant in other endophytes. Fouts et al. (2008) compared the genomes of 29 plant-associated bacteria and found only 45 "phytobacteria-specific" proteins, which dropped to 23 when they compared against a non-endophytic strain of *Klebsiella pneumoniae*. More recently, Levy et al. (2018) compared several thousand bacterial genomes to identify plant-associated gene patterns. They found that plant-associated bacteria—including endophytes but also ones from other plant compartments, like the rhizosphere—tend to have larger genomes, fewer mobile elements, and more genes related to carbohydrate metabolism. They also frequently encode protein domains that mimic plant

proteins; some of these are known to be involved in plant immunity, and many of the rest probably function in microbe–plant interaction and pathogenesis.

Although Levy et al. (2018) identified one intriguing instance of a pair of *Acidovorax* genes highly correlated with either commensalism or pathogenesis (called *Jekyll* and *Hyde*, respectively), most organisms do not seem to have such clear-cut differences. It seems likely that there are many possible routes to becoming an endophyte, and the differences between a given endophyte and its non-endophyte kin may be species-specific.

14.5.3 Genome-by-Genome Interactions

One of the most fascinating but least studied areas of maize–endophyte interaction is that of genome-by-genome interaction between host and endophytes. Genomic analyses of endophytic colonization are rare, and analyses of how genetic variation in the host or the endophyte affects their relationship are nearly nonexistent.

One of the few studies in this area found that shifts in maize small RNAs when infected with a bacterial endophyte downregulated a set of copper-related enzymes (Thiebaut et al. 2014). Copper is an important component in both plant defense pathways and lignin biosynthesis, and the authors speculate that manipulating either or both could make it easier for the endophyte to colonize the plant. More analyses along this vein are needed, especially for investigating the effect of maize genetic variation on endophytic colonization and impact. Although many authors mention that the effect of endophytes varies depending on the maize variety (e.g., Riggs et al. 2001; Montañez et al. 2008, 2012; Pan et al. 2008; Prischl et al. 2012; Naveed et al. 2014; Brusamarello-Santos et al. 2017), to our knowledge there have been no systematic efforts to identify the genetic basis of this variation.

Several authors have investigated if genetically modified maize hosts a different endophytic community than non-modified varieties. One would expect such differences to be minimal since none of the major genetic modifications in commercial maize are expected to impact microbes: The Cry toxin in Bt maize targets receptors in insect guts (Palma et al. 2014), while the various herbicide resistance genes simply allow the plant to continue normal metabolism in the presence of an herbicide. Empirical results largely confirm this expectation. For example, a test of a specific GFP-labeled strain of *Bacillus subtilis* found no difference between transgenic and control varieties (Sun et al. 2017). Another study found that the overall patterns of aboveground endophyte diversity were similar in Bt maize as in an isogenic non-Bt variety, though some specific bacterial clades (especially *Deinococcus* and *Pantoea*) were significantly different (Mashiane et al. 2017).

14.6 Manipulation of the Maize Endophyte Community

A major goal of endophyte research in maize and other crops is to understand how to manipulate microbial communities, including endophytes, to improve agricultural production and sustainability (Busby et al. 2017). Achieving that goal will require a much greater understanding of plant–microbe–environment interactions than we currently have, and for now it is probably safe to say that most manipulation of endophytes by growers is accidental. Some initial studies, however, have begun to shed light on how agricultural practices can affect endophytes.

14.6.1 Fertilizers

Fertilizer application can have a significant impact on maize endophyte communities, especially in the roots. Simply applying fertilizer to a field has been shown to reduce endophyte colonization and diversity (Roesch et al. 2006; Baldotto et al. 2012; Matsumura et al. 2015). Interestingly, one study that looked at both

endophyte RNA and DNA found that only the RNA pool was affected by fertilizer (Matsumura et al. 2015), implying it affected which endophytes were metabolically active but not which ones were present. Nitrogen fertilizer was also shown to decrease the association of nitrogen-fixing diazotrophs with maize early in the season but not later on (Roesch et al. 2006). These authors speculate that the difference was due to physiological changes in the plant and not the nitrogen supply per se. Applying organic versus mineral fertilizer has been shown to affect methanotrophic endophytes in the roots, possibly due to greater release of methane in the decaying organic fertilizer (Seghers et al. 2004).

14.6.2 Pesticides

The effect of pesticides on endophyte populations is almost entirely unknown. In at least one case, application of atrazine (a common preemergent herbicide) did not affect the resulting endophyte community (Seghers et al. 2004). Using several systemic fungicides in seed treatments affected the fungal endophytes in the leaves of soybean but not maize, possibly because of different fungicide mixes (Nettles et al. 2016). Even though glyphosate is currently the most broadly applied herbicide on the market (Benbrook 2016), there appear to be no peer-reviewed studies of its effect on the maize endosphere. The only report we could find is a single undergraduate thesis, which found no effect of glyphosate on bacteria in maize roots (Nolan 2016). This is surprising given that glyphosate is transported systemically throughout the plant and interferes with an enzyme present in most microbes (5-enolpyruvylshikimate-3-phosphate synthase, part of aromatic amino acid synthesis; Funke et al. 2006). Glyphosate is known to affect some members of the rhizosphere (Kremer and Means 2009), although others appear unaffected (Hart et al. 2009). Given their prevalence in large-scale agriculture, additional studies are needed to determine to what extent, if any, various pesticides alter the maize endosphere.

14.6.3 Biological Inoculation

The most straightforward way to manipulate the maize endophyte community is to directly infect maize plants. However, since many maize endophytes appear to come from the local environment (McInroy and Kloepper 1995a; Bokati et al. 2016), manipulations usually consist of inoculating seeds at the time of planting. In research settings, seeds are also usually surface-sterilized and planted into a sterile medium, but this approach will obviously not work for field conditions.

Artificial inoculation of endophytes has been shown to affect a range of plant health outcomes, including growth (Riggs et al. 2001; Bacon and Hinton 2002; Rosenblueth and Martinez-Romero 2004; Nassar et al. 2005; Hungria et al. 2010; Canellas et al. 2012; Montañez et al. 2012; Baldotto et al. 2012; Arruda et al. 2013; Knoth et al. 2013; Young et al. 2013; Akhtar et al. 2015), salinity tolerance (Akhtar et al. 2015), drought tolerance (Naveed et al. 2014), nutrient acquisition (Arruda et al. 2013; Young et al. 2013), disease resistance (Lee et al. 2009; Bacon and Hinton 2011; Chulze et al. 2015; Mousa et al. 2015), insect resistance (Lewis a et al. 1996), and yield (Hungria et al. 2010). Some endophytes have been particularly intensively studied for their beneficial effects, such members of the genus *Burkholderia* (Riggs et al. 2001; Arruda et al. 2013; Naveed et al. 2014; Akhtar et al. 2015) and *Herbaspirillum* (Riggs et al. 2001; Canellas et al. 2012; Baldotto et al. 2012; Arruda et al. 2013; do Amaral et al. 2014). The benefits from endophytes can be context-dependent, so that an endophyte that is beneficial under one set of circumstances can become neutral or even detrimental under another (e.g., increased fertilizer application; Riggs et al. 2001; Matsumura et al. 2015). Although possible mechanisms for endophyte-derived benefits are known (hormone synthesis, nitrogen fixation, competition with pathogens, etc.; Suman et al. 2016), in most cases the reasons why a specific endophyte affects maize in a given way are speculative at best.

Table 14.5 Example of maize biologicals with known or potential endophyte components

Manufacturer	Product	Known or potential maize endophytes
Advanced biological marketing	SabreEx	*Trichoderma*
Indigo Ag	Indigo Corn	Unknown[a]
Koppert	Panoramix	*Bacillus, Trichoderma*
Monsanto BioAg	Acceleron	*Penicillium*
Monsanto BioAg	QuickRoots	*Bacillus, Trichoderma*
MycoGold	Mycogold	*Azospirillum*
Nutri-Tech solutions	Nutri Life Platform	*Azospirillum, Bacillus, Pseudomonas, Streptomyces, Trichoderma*
TerraMax	MicroAZ	*Azospirillum*

[a]Product is known to include endophytes, but the exact species are not public

The lack of knowledge about how maize interacts with endophytes has not impeded their commercialization. "Biologicals"—a blanket term for agronomic inputs based on living organisms—are a rapidly growing market, with both small start-ups and major ag companies investing in them (see Zhang 2016). Many maize biologicals contain known or potential endophytes (Table 14.5), although in most cases their efficacy has not been independently verified. Given the current high rate of investment in biologicals, the number of endophyte-containing products for maize is likely to grow considerably over the next few years.

14.6.4 Other Management Impacts

Tillage, irrigation, crop rotation, and intercropping will presumably all have some effect on the maize endophyte community, either by altering the conditions for the plant or by altering the pool of organisms available to colonize it. Unfortunately, little data is available for how these affect the endophytic community in practice. Organic production apparently results in more diverse endophyte communities, at least among culturable bacteria (Xia et al. 2015). Tillage (Nolan 2016), pH (Adejumo and Orole 2010), and moisture content (Adejumo and Orole 2010) also affect endophytes, although the effects are variable and not always statistically significant.

14.7 Major Unsolved Questions in Maize Endophyte Research

A major conclusion from all this research is that what we do *not* know about maize endophytes vastly outweighs what we do. Even as companies rush to commercialize newly isolated strains, we still lack a basic understanding of how these organisms interact with their plant host, how they are affected by the environment and each other, and what contributions they make to maize phenotypes. Several major questions need to be resolved if we are to fully exploit maize endophytes for agriculture in the twenty-first century, including what constitutes a healthy endophyte community, what factors shape that community, and how can we harness endopyhtes for agriculture?

14.7.1 What Is a Healthy Maize Endophytic Community?

A broad definition of a "healthy" maize endophytic community would be one that increases the plant's fitness in a given environment. This

community would presumably protect the plant against diseases and pests, directly or indirectly improve the plant's nutrient status, and otherwise help the plant grow better than it would without the community. Although individual endophytes can sometimes be identified with some of these properties, the parameters of an entire healthy *community* are still unknown. A crucial aspect of this is that a healthy endophytic community is almost certainly context-dependent, and will most likely depend on the functions its members provide rather than what species they belong to.

14.7.2 How Do the Maize Plant, the Environment, and the Endophytes Themselves Shape the Maize Endophytic Community?

The maize endosphere is a miniature ecosystem, with all the complexity that implies. Founder effects, niche specialization, cooperation, competition, and many other ecological forces are all at play, complicated further by the environment itself—the maize plant—being its own living organism subject to macroscale interactions with its neighbors, herbivores, pests, humans, and the outside environment. Basic understanding of these individual forces is a necessary first step to understanding the rules of the maize endosphere. Integrating these rules into a systems-level understanding, complete with predictive models, will almost certainly be required to understand and exploit the full potential of maize endophytes.

14.7.3 How Can Maize Endophytes Be Harnessed to Improve Agriculture?

Maize is the most-produced grain on the planet, with over 1 billion metric tons harvested each year (Food and Agriculture Organization of the United Nations 2017). Given the growth and production benefits seen from individual endophytes, optimizing the entire maize endophytic community could have a significant impact on maize production and sustainability. The best we have right now, though, are individual endophytes that have been shown to improve maize growth under certain (often artificial) conditions. Intelligently deploying endophytes for agriculture will require a much better understanding of the conditions under which these endophytes do or do not work, how management practices interact with individual endophytes and the community as a whole, and what the potential benefits and limits of endophytes are in field settings.

14.8 Conclusion

The maize endosphere is a rich but poorly understood ecosystem. Endophytes can have significant impacts on their host plant, and understanding the rules of how endophytic communities assemble and how they impact plant health could be an important contribution to sustainable, integrated agriculture. Achieving this understanding will require better models of the interactions between microbes, plants, management, and the environment, including genomic analysis of both maize and its microbes. Integrating all these data into predictive models will be a significant challenge but has the potential to generate fundamental insights into maize–microbiome interaction with potentially wide-ranging impacts for global agriculture. Although endophytes will not solve all the challenges agriculture faces in the twenty-first century, they could become an important part of our efforts to achieve global sustainable production and food security.

References

Adejumo TO, Orole OO (2010) Effect of pH and moisture content on endophytic colonization of maize roots. Sci Res Essays 5:1655–1661

Akhtar SS, Andersen MN, Naveed M et al (2015) Interactive effect of biochar and plant growth-promoting bacterial endophytes on ameliorating salinity stress in maize. Funct Plant Biol 42:770

Ali S, Kandasamy S, Saldias S, Lazarovits G (2017) Corn and its interactions with bacterial communities. Rhizotrophs: plant growth promotion to bioremediation. Springer, Singapore, pp 145–163

Amin N (2013) Diversity of endophytic fungi from root of Maize var. Pulut (waxy corn local variety of South Sulawesi, Indonesia). Int J Curr Microbiol App Sci 2:148–154

Arnold AE, Mejía LC, Kyllo D et al (2003) Fungal endophytes limit pathogen damage in a tropical tree. Proc Natl Acad Sci USA 100:15649–15654

Arnold AE, Miadlikowska J, Higgins KL et al (2009) A phylogenetic estimation of trophic transition networks for ascomycetous fungi: are lichens cradles of symbiotrophic fungal diversification? Syst Biol 58:283–297

Arruda L, Beneduzi A, Martins A et al (2013) Screening of rhizobacteria isolated from maize (*Zea mays* L.) in Rio Grande do Sul State (South Brazil) and analysis of their potential to improve plant growth. Appl Soil Ecol 63:15–22

Bacon CW, Glenn AE, Yates IE (2008) Fusarium verticillioides: managing the endophytic association with maize for reduced fumonisins accumulation. Toxin Rev 27:411–446

Bacon CW, Hinton DM (2002) Endophytic and biological control potential of *Bacillus mojavensis* and related species. Biol Control 23:274–284

Bacon CW, Hinton DM (1996) Symptomless endophytic colonization of maize by *Fusarium moniliforme*. Can J Bot Can De Botanique 74:1195–1202

Bacon CW, Hinton DM (2011) *In planta* reduction of maize seedling stalk lesions by the bacterial endophyte *Bacillus mojavensis*. Can J Microbiol 57:485–492

Baldani JI, Baldani VLD, Seldin L, Döbereiner J (1986) Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a Root-Associated Nitrogen-Fixing Bacterium. Int J Syst Evol Microbiol 36:86–93

Baldotto MA, Borges Baldotto LE, Santana RB, Marciano CR (2012) Initial performance of maize in response to NPK fertilization combined with *Herbaspirillum seropedicae*. Revista Ceres 59

Ban Y, Xu Z, Yang Y et al (2017) Effect of dark septate endophytic fungus *Gaeumannomyces cylindrosporus* on plant growth, photosynthesis and Pb tolerance of maize (Zea mays L.). Pedosphere 27:283–292

Bary A (1866) Morphologie und physiologie der pilze. Flechten und myxomyceten. W Engelmann

Benbrook CM (2016) Trends in glyphosate herbicide use in the united states and globally. Environ Sci Eur 28:3

Besserer A, Puech-Pagès V, Kiefer P et al (2006) Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. PLoS Biol 4:e226

Bodhankar S, Grover M, Hemanth S et al (2017) Maize seed endophytic bacteria: dominance of antagonistic, lytic enzyme-producing Bacillus spp 3. Biotech 7:232

Bokati D, Herrera J, Poudel R (2016) Soil influences colonization of root-associated fungal endophyte communities of maize, wheat, and their progenitors. J Mycol Plant Pathol 2016:1–9

Brader G, Compant S, Vescio K et al (2017) ecology and genomic insights into plant-pathogenic and plant-nonpathogenic endophytes. Annu Rev Phytopathol 55:61–83

Brookes JJ (2017) Endophytes in maize (*Zea mays*) in New Zealand. Lincoln University

Brusamarello-Santos LC, Gilard F, Brule L et al (2017) Metabolic profiling of two maize (*Zea mays* L.) inbred lines inoculated with the nitrogen fixing plant-interacting bacteria *Herbaspirillum seropedicae* and *Azospirillum brasilense*. PLoS One 12:e0174576

Busby PE, Peay KG, Newcombe G (2016) Common foliar fungi of *Populus trichocarpa* modify *Melampsora* rust disease severity. New Phytol 209:1681–1692

Busby PE, Soman C, Wagner MR et al (2017) Research priorities for harnessing plant microbiomes in sustainable agriculture. PLoS Biol 15

Canellas LP, Balmori DM, Médici LO et al (2012) A combination of humic substances and *Herbaspirillum seropedicae* inoculation enhances the growth of maize (Zea mays L.). Plant Soil 366:119–132

Casanovas EM, Barassi CA, Sueldo RJ (2002) *azospirillum* inoculation mitigates water stress effects in maize seedlings. Cereal Res Commun 30:343–350

Chelius MK, Triplett EW (2001) The diversity of Archaea and Bacteria in association with the roots of *Zea mays* L. Microb Ecol 41:252–263

Chelius MK, Triplett EW (2000) Immunolocalization of dinitrogenase reductase produced by *Klebsiella pneumoniae* in association with *Zea mays* L. Appl Environ Microbiol 66:783–787

Chelius MK, Triplett EW, Others (2000) Diazotrophic endophytes associated with maize. Prokaryotic nitrogen fixation: a model system for the analysis of a biological process, pp 779–791

Chen L, Xin X, Zhang J et al (2017) Soil characteristics overwhelm cultivar effects on the structure and assembly of root-associated microbiomes of modern maize. Pedosphere. https://doi.org/10.1016/S1002-0160(17)60370-9

Cherry AJ, Banito A, Djegui D, Lomer C (2004) Suppression of the stem-borer *Sesamia calamistis* (Lepidoptera; Noctuidae) in maize following seed dressing, topical application and stem injection with African isolates of *Beauveria bassiana*. Int J Pest Manage 50:67–73

Cherry AJ, Lomer CJ, Djegui D, Schulthess F (1999) Pathogen incidence and their potential as microbial control agents in IPM of maize stem borers in West Africa. Biocontrol 44:301–327

Chulze SN, Palazzini JM, Torres AM et al (2015) Biological control as a strategy to reduce the impact of mycotoxins in peanuts, grapes and cereals in Argentina. Food Addit Contam Part A Chem Anal Control Expo Risk Assess 32:471–479

Cohen AC, Travaglia CN, Bottini R, Piccoli PN (2009) Participation of abscisic acid and gibberellins produced by endophytic *Azospirillum* in the alleviation of drought effects in maize. Botany 87:455–462

Compant S, Clément C, Sessitsch A (2010) Plant growth-promoting bacteria in the rhizo- and endosphere of plants: their role, colonization, mechanisms involved and prospects for utilization. Soil Biol Biochem 42:669–678

Contreras-Cornejo HA, Macías-Rodríguez L, del-Val E, Larsen J (2017) The root endophytic fungus *Trichoderma atroviride* induces foliar herbivory resistance in maize plants. Appl Soil Ecol. https://doi.org/10.1016/j.apsoil.2017.10.004

da Silva DAF, Cotta SR, Vollú RE et al (2014) Endophytic microbial community in two transgenic maize genotypes and in their near-isogenic non-transgenic maize genotype. BMC Microbiol 14:332

David AS, Seabloom EW, May G (2016) Plant host species and geographic distance affect the structure of aboveground fungal symbiont communities, and environmental filtering affects belowground communities in a coastal dune ecosystem. Microb Ecol 71:912–926

de Araujo JM, da Silva AC, Azevedo JL (2000) Isolation of endophytic actinomycetes from roots and leaves of maize (Zea may L.). Braz Arch Biol Technol 43:447–451

Ding T, Su B, Chen X et al (2017) an endophytic bacterial strain isolated from *Eucommia ulmoides* inhibits southern corn leaf blight. Front Microbiol 8:903

do Amaral FP, Bueno JCF, Hermes VS, Arisi ACM (2014) Gene expression analysis of maize seedlings (DKB240 variety) inoculated with plant growth promoting bacterium *Herbaspirillum seropedicae*. Symbiosis 62:41–50

ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. Science 306:636–640

Estrada P, Mavingui P, Cournoyer B et al (2002) A N2-fixing endophytic *Burkholderia* sp. associated with maize plants cultivated in Mexico. Can J Microbiol 48:285–294

Ettinger CL, Shehata HR, Johnston-Monje D et al (2015) Draft genome sequence of *Burkholderia gladioli* strain UCD-UG_CHAPALOTE (Phylum Proteobacteria). Genome Announc 3. https://doi.org/10.1128/genomea.01462-14

Ferrari CS, Amaral FP, Bueno JCF et al (2014) Expressed proteins of *Herbaspirillum seropedicae* in maize (DKB240) roots-bacteria interaction revealed using proteomics. Appl Biochem Biotechnol 174:2267–2277

Figueiredo JEF, Gomes EA, Guimarães CT et al (2009) Molecular analysis of endophytic bacteria from the genus *Bacillus* isolated from tropical maize (*Zea mays* L.). Braz J Microbiol 40:522–534

Fisher PJ, Petrini O, Scott HML (1992) The distribution of some fungal and bacterial endophytes in maize (*Zea-Mays* L). New Phytol 122:299–305

Flandrois J-P, Perrière G, Gouy M (2015) leBIBIQBPP: a set of databases and a webtool for automatic phylogenetic analysis of prokaryotic sequences. BMC Bioinform 16:251

Food and Agriculture Organization of the United Nations (2017) FAOSTAT statistics database

Fouts DE, Tyler HL, DeBoy RT et al (2008) Complete genome sequence of the N2-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. PLoS Genet 4:e1000141

Frank AC (2011) The genomes of endophytic bacteria. Endophytes of forest trees. Springer, Dordrecht, pp 107–136

Funke T, Han H, Healy-Fried ML et al (2006) Molecular basis for the herbicide resistance of Roundup Ready crops. Proc Natl Acad Sci U S A 103:13010–13015

Gams W (1971) Cephalosporium-artige schimmelpilze (Hyphomycetes). Gustav Fischer Verlag, Stuttgart, West Germany, p 262

Gao J-L, Sun P, Wang X-M et al (2017) *Microbacterium zeae* sp. nov., an endophytic bacterium isolated from maize stem. Antonie Van Leeuwenhoek 110:697–704

Gao Z, Zhuang J, Chen J et al (2004) Population of entophytic bacteria in maize roots and its dynamic analysis. Ying Yong Sheng Tai Xue Bao 15:1344–1348

Giauque H, Hawkes CV (2013) Climate affects symbiotic fungal endophyte diversity and performance. Am J Bot 100:1435–1444

Glenn AE, Gold SE, Bacon CW (2002) Fdb1 and Fdb2, *Fusarium verticillioides* loci necessary for detoxification of preformed antimicrobials from corn. Mol Plant Microbe Interact 15:91–101

Glenn AE, Hinton DM, Yates IE, Bacon CW (2001) Detoxification of corn antimicrobial compounds as the basis for isolating *Fusarium verticillioides* and some other *Fusarium* species from corn. Appl Environ Microbiol 67:2973–2981

Gold SE, Blacutt AA, Meinersmann RJ, Bacon CW (2014) Whole-genome shotgun sequence of *Bacillus mojavensis* strain RRC101, an endophytic bacterium antagonistic to the mycotoxigenic endophytic fungus *Fusarium verticillioides*. Genome Announc 2. https://doi.org/10.1128/genomea.01090-14

Gond SK, Bergen MS, Torres MS, White JF Jr (2015) Endophytic *Bacillus* spp. Produce antifungal lipopeptides and induce host defence gene expression in maize. Microbiol Res 172:79–87

Goodfellow P (1995) A big book of the human genome. Complement Endeav Nat 377:285–286

Gutiérrez-Zamora ML, Martínez-Romero E (2001) Natural endophytic association between Rhizobium etli and maize (*Zea mays* L.). J Biotechnol 91:117–126

Gyaneshwar P, James EK, Reddy PM, Ladha JK (2002) *Herbaspirillum* colonization increases growth and nitrogen accumulation in aluminium-tolerant rice varieties. New Phytol 154:131–145

Hardoim PR, van Overbeek LS, Berg G et al (2015) The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. Microbiol Mol Biol Rev 79:293–320

Harman GE, Petzoldt R, Comis A, Chen J (2004) Interactions between *Trichoderma harzianum* strain T22 and maize inbred line Mo17 and effects of these interactions on diseases caused by *Pythium ultimum* and *Colletotrichum graminicola*. Phytopathology 94:147–153

Hart MM, Powell JR, Gulden RH et al (2009) Separating the effect of crop from herbicide on soil microbial communities in glyphosate-resistant corn. Pedobiologia 52:253–262

Hawkes CV, Connor EW (2017) Translating phytobiomes from theory to practice: ecological and evolutionary considerations. Phytobiomes 1:57–69

Helgason T, Fitter AH (2009) Natural selection and the evolutionary ecology of the arbuscular mycorrhizal fungi (Phylum Glomeromycota). J Exp Bot 60:2465–2480

He Y, Yang Z, Li M et al (2017) Effects of a dark septate endophyte (DSE) on growth, cadmium content, and physiology in maize under cadmium stress. Environ Sci Pollut Res Int 24:18494–18504

Hinton DM, Bacon CW (1995) *Enterobacter cloacae* is an endophytic symbiont of corn. Mycopathologia 129:117–125

Hodgson S, de Cates C, Hodgson J et al (2014) Vertical transmission of fungal endophytes is widespread in forbs. Ecol Evol 4:1199–1208

Hollowell AC, Regus JU, Turissini D et al (2016) Metapopulation dominance and genomic-island acquisition of *Bradyrhizobium* with superior catabolic capabilities. Proc Biol Sci 283. https://doi.org/10.1098/rspb.2016.0496

Hungria M, Campo RJ, Souza EM, Pedrosa FO (2010) Inoculation with selected strains of *Azospirillum brasilense* and *A. lipoferum* improves yields of maize and wheat in Brazil. Plant Soil 331:413–425

Ikeda AC, Bassani LL, Adamoski D et al (2013) Morphological and genetic characterization of endophytic bacteria isolated from roots of different maize genotypes. Microb Ecol 65:154–160

Iniguez AL, Dong Y, Triplett EW (2004) Nitrogen fixation in wheat provided by Klebsiella pneumoniae 342. Mol Plant Microbe Interact 17:1078–1085

Jiao Y, Peluso P, Shi J et al (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527

Johnston-Monje D, Lundberg DS, Lazarovits G et al (2016) Bacterial populations in juvenile maize rhizospheres originate from both seed and soil. Plant Soil 405:337–355

Johnston-Monje D, Mousa WK, Lazarovits G, Raizada MN (2014) Impact of swapping soils on the endophytic bacterial communities of pre-domesticated, ancient and modern maize. BMC Plant Biol 14

Johnston-Monje D, Raizada MN (2011) Conservation and diversity of seed associated endophytes in *Zea* across boundaries of evolution, ethnography and ecology. PLoS One 6:e20396

Knoth JL, Kim S-H, Ettl GJ, Doty SL (2013) Effects of cross host species inoculation of nitrogen-fixing endophytes on growth and leaf physiology of maize. GCB Bioenerg 5:408–418

Krause A, Ramakumar A, Bartels D et al (2006) Complete genome of the mutualistic, N2-fixing grass endophyte *Azoarcus* sp. Strain BH72. Nat Biotechnol 24:1385–1391

Kremer RJ, Means NE (2009) Glyphosate and glyphosate-resistant crop interactions with rhizosphere microorganisms. Eur J Agron 31:153–161

Kroll S, Agler MT, Kemen E (2017) Genomic dissection of host–microbe and microbe–microbe interactions for advanced plant breeding. Curr Opin Plant Biol 36:71–78

Kuldau GA, Yates IE (2000) Evidence for *Fusarium* endophytes in cultivated and wild plants. Microb Endophytes 85–117

Lamb TG, Tonkyn DW, Kluepfel DA (1996) Movement of *Pseudomonas aureofaciens* from the rhizosphere to aerial plant tissue. Can J Microbiol 42:1112–1120

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Lee K, Pan JJ, May G (2009) Endophytic *Fusarium verticillioides* reduces disease severity caused by *Ustilago maydis* on maize. FEMS Microbiol Lett 299:31–37

Leggett HC, Buckling A, Long GH, Boots M (2013) Generalism and the evolution of parasite virulence. Trends Ecol Evol 28:592–596

Leslie JF, Pearson CAS, Nelson PE, Toussoun TA (1990) *Fusarium* spp. From corn, sorghum, and soybean fields in the central and eastern United States. Ecol Stud 44:66

Levy A, Salas Gonzalez I, Mittelviefhaus M et al (2018) Genomic features of bacterial adaptation to plants. Nat Genet 50:138–150

Lewis a LC, Berry b EC, Obrycki c JJ, Bing c LA (1996) Aptness of insecticides (*Bacillus thuringiensis* and *carbofuran*) with endophytic *Beauveria bassiana*, in suppressing larval populations of the European corn borer. ELSEVIER Agric Ecosyst Environ 57:27–34

Lewis LC, Cossentine JE (1986) Season long intraplant epizootics of entomopathogens, *Beauveria bassiana* and *Nosema pyrausta*, in a corn agroecosystem. Biocontrol 31:363–369

Li T, Liu MJ, Zhang XT et al (2011) Improved tolerance of maize (*Zea mays* L.) to heavy metals by colonization of a dark septate endophyte (DSE) *Exophiala pisciphila*. Sci Total Environ 409:1069–1074

Liu X, Hao L, Li D et al (2015a) Long non-coding RNAs and their biological roles in plants. Genomics Proteomics Bioinform 13:137–147

Liu Y, Wang R, Cao Y et al (2015b) Identification and antagonistic activity of endophytic bacterial strain *Paenibacillus* sp. 5 L8 isolated from the seeds of maize (*Zea mays* L., Jingke 968). Ann Microbiol 66:653–660

Liu X, Zhao H, Chen S (2006) Colonization of maize and rice plants by strain *Bacillus megaterium* C4. Curr Microbiol 52:186–190

Liu Y, Wang R, Li Y et al (2017) High-throughput sequencing-based analysis of the composition and diversity of endophytic bacterial community in seeds of "Beijing" hybrid maize planted in China. Plant Growth Regul 81:317–324

Liu Y, Zuo S, Xu L et al (2012a) Study on diversity of endophytic bacterial communities in seeds of hybrid maize and their parental lines. Arch Microbiol 194:1001–1012

Liu Y, Zuo S, Zou Y et al (2012b) Investigation on diversity and population succession dynamics of endophytic bacteria from seeds of maize (*Zea mays* L., Nongda108) at different growth stages. Ann Microbiol 63:71–79

López-Guerrero MG, Ormeño-Orrillo E, Velázquez E et al (2012) *Rhizobium etli* taxonomy revised with novel genomic data and analyses. Syst Appl Microbiol 35:353–358

Lucangeli C, Bottini R (1997) Effects of *Azospirillum* spp. On endogenous gibberellin content and growth of maize (*Zea mays* L.) treated with uniconazole. Symbiosis 23:63–71

Ludueña LM, Anzuay MS, Angelini JG et al (2018) Strain Serratia sp. S119: a potential biofertilizer for peanut and maize and a model bacterium to study phosphate solubilization mechanisms. Appl Soil Ecol. https://doi.org/10.1016/j.apsoil.2017.12.024

Ludueña LM, Anzuay MS, Angelini JG et al (2017) Role of bacterial pyrroloquinoline quinone in phosphate solubilizing ability and in plant growth promotion on strain Serratia sp. S119. Symbiosis 72:31–43

Madmony A, Chernin L, Pleban S et al (2005) *Enterobacter cloacae*, an obligatory endophyte of pollen grains of Mediterranean pines. Folia Microbiol 50:209–216

Mano H, Morisaki H (2008) Endophytic bacteria in the rice plant. Microbes Environ 23:109–117

Mashiane RA, Ezeokoli OT, Adeleke RA, Bezuidenhout CC (2017) Metagenomic analyses of bacterial endophytes associated with the phyllosphere of a Bt maize cultivar and its isogenic parental line from South Africa. World J Microbiol Biotechnol 33:80

Matsumura EE, Secco VA, Moreira RS et al (2015) Composition and activity of endophytic bacterial communities in field-grown maize plants inoculated with *Azospirillum brasilense*. Ann Microbiol 65:2187–2200

McInroy JA, Kloepper JW (1991) Endophytic bacteria from field-grown corn and cotton

McInroy JA, Kloepper JW (1995a) Population dynamics of endophytic bacteria in field-grown sweet corn and cotton. Can J Microbiol 41:895–901

McInroy JA, Kloepper JW (1995b) Survey of indigenous bacterial endophytes from cotton and sweet corn. Plant Soil 173:337–342

Medini D, Donati C, Tettelin H et al (2005) The microbial pan-genome. Curr Opin Genet Dev 15:589–594

Menéndez E, Ramirez-Bahena MH, Peix A et al (2016) Analysis of cultivable endophytic bacteria in roots of maize in a soil from león province in mainland spain. Biological nitrogen fixation and beneficial plant-microbe interaction. Springer, Cham, pp 45–53

Messing J, Bharti AK, Karlowski WM et al (2004) Sequence composition and genome organization of maize. Proc Natl Acad Sci USA 101:14349–14354

Michael TP, Jackson S (2013) The first 50 plant genomes. Plant Genome 6

Mitter B, Petric A, Shin MW et al (2013) Comparative genome analysis of *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic lifestyles based on interaction strategies with host plants. Front Plant Sci 4:120

Mohanty SR, Dubey G, Kollah B (2017) Endophytes of *Jatropha curcas* promote growth of maize. Rhizosphere 3:20–28

Montañez A, Abreu C, Gill PR et al (2008) Biological nitrogen fixation in maize (*Zea mays* L.) by 15 N isotope-dilution and identification of associated culturable diazotrophs. Biol Fertil Soils 45:253–263

Montañez A, Blanco AR, Barlocco C et al (2012) Characterization of cultivable putative endophytic plant growth promoting bacteria associated with maize cultivars (*Zea mays* L.) and their inoculation effects in vitro. Appl Soil Ecol 58:21–28

Monteiro RA, Schmidt MA, de Baura VA et al (2008) Early colonization pattern of maize (*Zea mays* L. Poales, Poaceae) roots by *Herbaspirillum seropedicae* (Burkholderiales, Oxalobacteraceae). Genet Mol Biol 31:932–937

Moran NA, Sloan DB (2015) The hologenome concept: helpful or hollow? PLoS Biol 13

Mousa WK, Shearer CR, Limay-Rios V et al (2015) Bacterial endophytes from wild maize suppress *Fusarium graminearum* in modern maize and inhibit mycotoxin accumulation. Front Plant Sci 6:805

Nassar AH, El-Tarabily KA, Sivasithamparam K (2005) Promotion of plant growth by an auxin-producing isolate of the yeast *Williopsis saturnus* endophytic in maize (*Zea mays* L.) roots. Biol Fertil Soils 42:97–108

Naveed M, Mitter B, Reichenauer TG et al (2014) Increased drought stress resilience of maize through endophytic colonization by *Burkholderia phytofirmans* PsJN and *Enterobacter* sp. FD17. Environ Exp Bot 97:30–39

Nelson PG, May G (2017) Coevolution between Mutualists and parasites in symbiotic communities may lead to the evolution of lower virulence. Am Nat 190:803–817

Nettles R, Watkins J, Ricks K, et al (2016) Influence of pesticide seed treatments on rhizosphere fungal and bacterial communities and leaf fungal endophyte communities in maize and soybean. Appl Soil Ecol 102:61–69

Nolan BLF (2016) The effects of tillage, glyphosate, and genetic modification on bacterial root endophyte composition in *Zea mays*. Undergraduate, The University of Mississippi

Orole OO, Adejumo TO (2009) Activity of fungal endophytes against four maize wilt pathogens. Afr J Microbiol Res 3:969–973

Orole OO, Adejumo TO (2011) Bacterial and fungal endophytes associated with grains and roots of maize. J Ecol Nat Environ 3:298–303

Palma L, Muñoz D, Berry C et al (2014) *Bacillus thuringiensis toxins*: an overview of their biocidal activity. Toxins 6:3296–3325

Palus JA, Borneman J, Ludden PW, Triplett EW (1996) A diazotrophic bacterial endophyte isolated from stems of *Zea mays* L. and *Zea luxurians* Iltis and Doebley. Plant Soil 186:135–142

Pan JJ, Baumgarten AM, May G (2008) Effects of host plant environment and *Ustilago maydis* infection on the fungal endophyte community of maize (*Zea mays*). New Phytol 178:147–156

Pan JJ, May G (2009) Fungal-fungal associations affect the assembly of endophyte communities in maize (*Zea mays*). Microb Ecol 58:668–678

Pedrosa FO, Monteiro RA, Wassem R et al (2011) Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses. PLoS Genet 7:e1002064

Pereira P, Ibáñez F, Rosenblueth M, et al (2011) Analysis of the bacterial diversity associated with the roots of maize (*Zea mays* L.) through culture-dependent and culture-independent methods. ISRN Ecol 2011

Pertea M, Salzberg SL (2010) Between a chicken and a grape: estimating the number of human genes. Genome Biol 11:206

Peterson SW, Jurjević Ž (2017) New species of *Talaromyces* isolated from maize, indoor air, and other substrates. Mycologia 109:537–556

Petrini O (1991) Fungal Endophytes of Tree Leaves. Microbial ecology of leaves. Springer, New York, NY, pp 179–197

Poling SM, Wicklow DT, Rogers KD, Gloer JB (2008) *Acremonium zeae*, a protective endophyte of maize, produces dihydroresorcylide and 7-hydroxydihydroresorcylides. J Agric Food Chem 56:3006–3009

Potshangbam M, Devi SI, Sahoo D, Strobel GA (2017) Functional Characterization of Endophytic Fungal Community Associated with Oryza sativa L and *Zea mays*L. Front Microbiol 8:325

Pray L (2008) Eukaryotic genome complexity. Nature. Education 1:96

Prischl M, Hackl E, Pastar M et al (2012) Genetically modified Bt maize lines containing cry3Bb1, cry1A105 or cry1Ab2 do not affect the structure and functioning of root-associated endophyte communities. Appl Soil Ecol 54:39–48

Proença DN, Schwab S, Baldani JI, Morais PV (2017) Diversity and function of endophytic microbial community of plants with economical potential. In: de Azevedo JL, Quecine MC (eds) Diversity and benefits of microorganisms from the tropics. Springer International Publishing, Cham, pp 209–243

Reinhold-Hurek B, Hurek T (2011) Living inside plants: bacterial endophytes. Curr Opin Plant Biol 14:435–443

Riggs PJ, Chelius MK, Iniguez AL et al (2001) Enhanced maize productivity by inoculation with diazotrophic bacteria. Aust J Plant Physiol 28:829–836

Rijavec T, Lapanje A, Dermastia M, Rupnik M (2007) Isolation of bacterial endophytes from germinated maize kernels. Can J Microbiol 53:802–808

Rodriguez Estrada AE, Hegeman A, Corby Kistler H, May G (2011) In vitro interactions between *Fusarium verticillioides* and *Ustilago maydis* through real-time PCR and metabolic profiling. Fungal Genet Biol 48:874–885

Roesch LFW, Camargo FAO, Bento FM, Triplett EW (2008) Biodiversity of diazotrophic bacteria within the soil, root and stem of field-grown maize. Plant Soil 302:91–104

Roesch LFW, Olivares FL, Pereira Passaglia LM et al (2006) Characterization of diazotrophic bacteria associated with maize: effect of plant genotype, ontogeny and nitrogen-supply. World J Microbiol Biotechnol 22:967–974

Roncato-Maccari LDB, Ramos HJO, Pedrosa FO et al (2003) Endophytic *Herbaspirillum seropedicae* expresses nif genes in gramineous plants. FEMS Microbiol Ecol 45:39–47

Roos IMM, Hattingh MJ (1983) Scanning electron microscopy of *Pseudomonas syringae* pv, morsprunorum on sweet cherry leaves. J Phytopathol 108:18–25

Rosenblueth M, Martinez-Romero E (2004) *Rhizobium etli* maize populations and their competitiveness for root colonization. Arch Microbiol 181:337–344

Sandhya V, Shrivastava M, Ali SZ, Sai Shiva V (2017) Endophytes from maize with plant growth promotion and biocontrol activity under drought stress. Russ Agric Sci 43:22–34

Santi C, Bogusz D, Franche C (2013) Biological nitrogen fixation in non-legume plants. Ann Bot 111:743–767

Santos F, Peñaflor MFGV, Paré PW et al (2014) A novel interaction between plant-beneficial rhizobacteria and roots: colonization induces corn resistance against the root herbivore Diabrotica speciosa. PLoS One 9: e113280

Saunders M, Glenn AE, Kohn LM (2010) Exploring the evolutionary ecology of fungal endophytes in agricultural systems: using functional traits to reveal mechanisms in community processes. Evol Appl 3:525–537

Saunders M, Kohn LM (2009) Evidence for alteration of fungal endophyte community assembly by host defense compounds. New Phytol 182:229–238

Saunders M, Kohn LM (2008) Host-synthesized secondary compounds influence the in vitro interactions between fungal endophytes of maize. Appl Environ Microbiol 74:136–142

Schardl CL (2001) *Epichloë festucae* and related mutualistic symbionts of grasses. Fungal Genet Biol 33:69–82

Seghers D, Wittebolle L, Top EM et al (2004) Impact of agricultural practices on the *Zea mays* L. Endophytic community. Appl Environ Microbiol 70:1475–1482

Sessitsch A, Coenye T, Sturz AV et al (2005) *Burkholderia phytofirmans* sp. nov, a novel plant-associated bacterium with plant-beneficial properties. Int J Syst Evol Microbiol 55:1187–1192

Shehata H (2016) Molecular and physiological mechanisms underlying the antifungal and nutrient acquisition activities of beneficial microbes

Shehata HR, Griffiths MW, Raizada MN (2017) seeds of the wild progenitor of maize possess bacteria that antagonize foodborne pathogens. Foodborne Pathog Dis 14:202–209

Shehata HR, Lyons EM, Jordan KS, Raizada MN (2016) Bacterial endophytes from wild and ancient maize are able to suppress the fungal pathogen *Sclerotinia homoeocarpa*. J Appl Microbiol 120:756–769

Shehata HR, Raizada MN (2017) A *Burkholderia* endophyte of the ancient maize landrace Chapalote utilizes c-di-GMP-dependent and independent signaling to suppress diverse plant fungal pathogen targets. FEMS Microbiol Lett 364. https://doi.org/10.1093/femsle/fnx138

Sheibani-Tezerji R, Naveed M, Jehl M-A et al (2015) The genomes of closely related *Pantoea ananatis* maize seed endophytes having different effects on the host plant differ in secretion system genes and mobile genetic elements. Front Microbiol 6:440

Siegel MR, Johnson MC, Varney DR et al (1984) A fungal endophyte in tall fescue: incidence and dissemination. Phytopathology 74:932–937

Sobowale AA, Cardwell KF, Odebode AC et al (2007) Persistence of *Trichoderma* species within maize stem against *Fusarium verticillioides*. Arch Phytopathol Plant Prot 40:215–231

Suman A, Yadav AN, Verma P (2016) Endophytic microbes in crops: diversity and beneficial impact for sustainable agriculture. Microbial inoculants in sustainable agricultural productivity. Springer, New Delhi, pp 117–143

Sun C, Geng L, Wang M, et al (2017) No adverse effects of transgenic maize on population dynamics of endophytic *Bacillus subtilis* strain B916-gfp. Microbiologyopen 6. https://doi.org/10.1002/mbo3.404

Syed NH, Kalyna M, Marquez Y et al (2012) Alternative splicing in plants–coming of age. Trends Plant Sci 17:616–623

Szilagyi-Zecchin VJ, Ikeda AC, Hungria M, et al (2014) Identification and characterization of endophytic bacteria from corn (*Zea mays* L.) roots with biotechnological potential in agriculture. AMB Express 4:26

Takenaka M, Zehrmann A, Verbitskiy D et al (2013) RNA editing in plants and its evolution. Annu Rev Genet 47:335–352

Taurian T, Anzuay MS, Angelini JG et al (2010) Phosphate-solubilizing peanut associated bacteria: screening for plant growth-promoting activities. Plant Soil 329:421–431

Thiebaut F, Rojas CA, Grativol C et al (2014) Genome-wide identification of microRNA and siRNA responsive to endophytic beneficial diazotrophic bacteria in maize. BMC Genom 15:766

Toruño TY, Stergiopoulos I, Coaker G (2016) Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. Annu Rev Phytopathol 54:419–441

U'Ren JM, Lutzoni F, Miadlikowska J et al (2012) Host and geographic structure of endophytic and endolichenic fungi at a continental scale. Am J Bot 99:898–914

Van Wees SCM, Van der Ent S, Pieterse CMJ (2008) Plant immune responses triggered by beneficial microbes. Curr Opin Plant Biol 11:443–448

Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351

Vurukonda SSKP, Vardharajula S, Shrivastava M, SkZ A (2016) Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. Microbiol Res 184:13–24

Wagner BL, Lewis LC (2000) Colonization of corn, *Zea mays*, by the entomopathogenic fungus *Beauveria bassiana*. Appl Environ Microbiol 66:3468–3473

Wang J-L, Li T, Liu G-Y et al (2016) Unraveling the role of dark septate endophyte (DSE) colonizing maize (*Zea mays*) under cadmium stress: physiological, cytological and genic aspects. Sci Rep 6:22028

Weilharter A, Mitter B, Shin MV et al (2011) Complete genome sequence of the plant growth-promoting endophyte *Burkholderia phytofirmans* strain PsJN. J Bacteriol 193:3383–3384

Wicklow DT, Poling SM (2009) Antimicrobial activity of pyrrocidines from *Acremonium zeae* against endophytes and pathogens of maize. Phytopathology 99:109–115

Wicklow DT, Roth S, Deyrup ST, Gloer JB (2005) A protective endophyte of maize: *Acremonium zeae* antibiotics inhibitory to *Aspergillus flavus* and *Fusarium verticillioides*. Mycol Res 109:610–618

Williams WP, Paul Williams W, Ozkan S et al (2011) Ear rot, aflatoxin accumulation, and fungal biomass in maize after inoculation with *Aspergillus flavus*. Field Crops Res 120:196–200

Xia Y, DeBolt S, Dreyer J et al (2015) Characterization of culturable bacterial endophytes and their capacity to promote plant growth from plants grown using organic or conventional practices. Front Plant Sci 6:490

Yang X, Xie L, Li Y, Wei C (2009) More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. PLoS One 4:e6074

Young L-S, Hameed A, Peng S-Y et al (2013) Endophytic establishment of the soil isolate *Burkholderia* sp. CC-Al74 enhances growth and P-utilization rate in maize (*Zea mays* L.). Appl Soil Ecol 66:40–47

Zhang S (2016) Good riddance, chemicals: microbes are farming's hot new pesticides. https://www.wired.com/2016/03/good-riddance-chemicals-microbes-farmings-hot-new-pesticides/

Zimmerman NB, Vitousek PM (2012) Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. Proc Natl Acad Sci USA 109:13022–13027

Zinniel DK, Lambrecht P, Harris NB et al (2002) Isolation and characterization of endophytic colonizing bacteria from agronomic crops and prairie plants. Appl Environ Microbiol 68:2198–2208

Transcriptomic Dissection of Maize Root System Development

15

Peng Yu, Caroline Marcon, Jutta A. Baldauf, Felix Frey, Marcel Baer and Frank Hochholdinger

Abstract

RNA-Seq (RNA-Sequencing) allows for precise quantitative determination of global gene expression patterns and has therefore revolutionized transcriptome analyses in maize. In recent years, genetic analyses have identified numerous genes that control maize root system architecture and root hair elongation. In addition, RNA-Seq has been applied to dissect structure and function of individual roots. In this chapter, we summarize the current state of the transcriptomic dissection of maize root development on the level of whole roots, tissues, and individual cells. Moreover, we highlight the current knowledge of transcriptome responses of maize roots to drought stress and nutrient availability. Finally, we outline novel findings related to gene expression plasticity in primary roots of maize hybrids during the early manifestation of heterosis.

15.1 Introduction

Genome sequencing of the maize inbred line B73 was a major landmark in maize genetics (Schnable et al. 2009). The recent *de novo* assembly and annotation of the maize B73 genome sequence led to the identification of 39,324 protein-coding genes (https://www.maizegdb.org/assembly, v4). Nevertheless, the function of most of these genes remains elusive (Jiao et al. 2017). Functional genomic tools such as RNA-Seq (RNA-Sequencing) are critical to explore the transcriptomic landscape during plant development, tissue patterning, and cell organization. Maize displays a complex root system architecture (Fig. 15.1) as illustrated by the presence of different embryonic and post-embryonic root types (reviewed in Yu et al. 2016; Hochholdinger et al. 2018a). This chapter summarizes the current progress in understanding the transcriptomic landscape controlling the function of the maize root system and its responses to environmental stimuli. Moreover, we highlight the transcriptomic plasticity of maize hybrids compared to their parental inbred lines during the early manifestation of heterosis. A review summarizing the status of the proteomic dissection of the maize root system has been published recently (Hochholdinger et al. 2018b).

P. Yu · C. Marcon · J. A. Baldauf · F. Frey
M. Baer · F. Hochholdinger (✉)
INRES, Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany
e-mail: hochholdinger@uni-bonn.de
URL: https://www.cfg.uni-bonn.de

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_15

15.2 Transcriptomic Dissection of Different Root Types, Tissues, and Cell Types

15.2.1 Root Type-Specific Transcriptomics

The rootstock of maize consists of an embryonically formed primary root and a variable number of embryonic seminal roots and post-embryonic shoot-borne crown and brace roots (Fig. 15.1; reviewed in Hochholdinger et al. 2004a). In the first weeks after germination, primary and seminal roots make up the major portion of the seedling rootstock (Hochholdinger et al. 2004b) and are thus vital for the early vigor of young maize seedlings (Peter et al. 2009). Shoot-borne roots formed from consecutive shoot nodes below the soil level are designated crown roots, while their counterparts initiated from

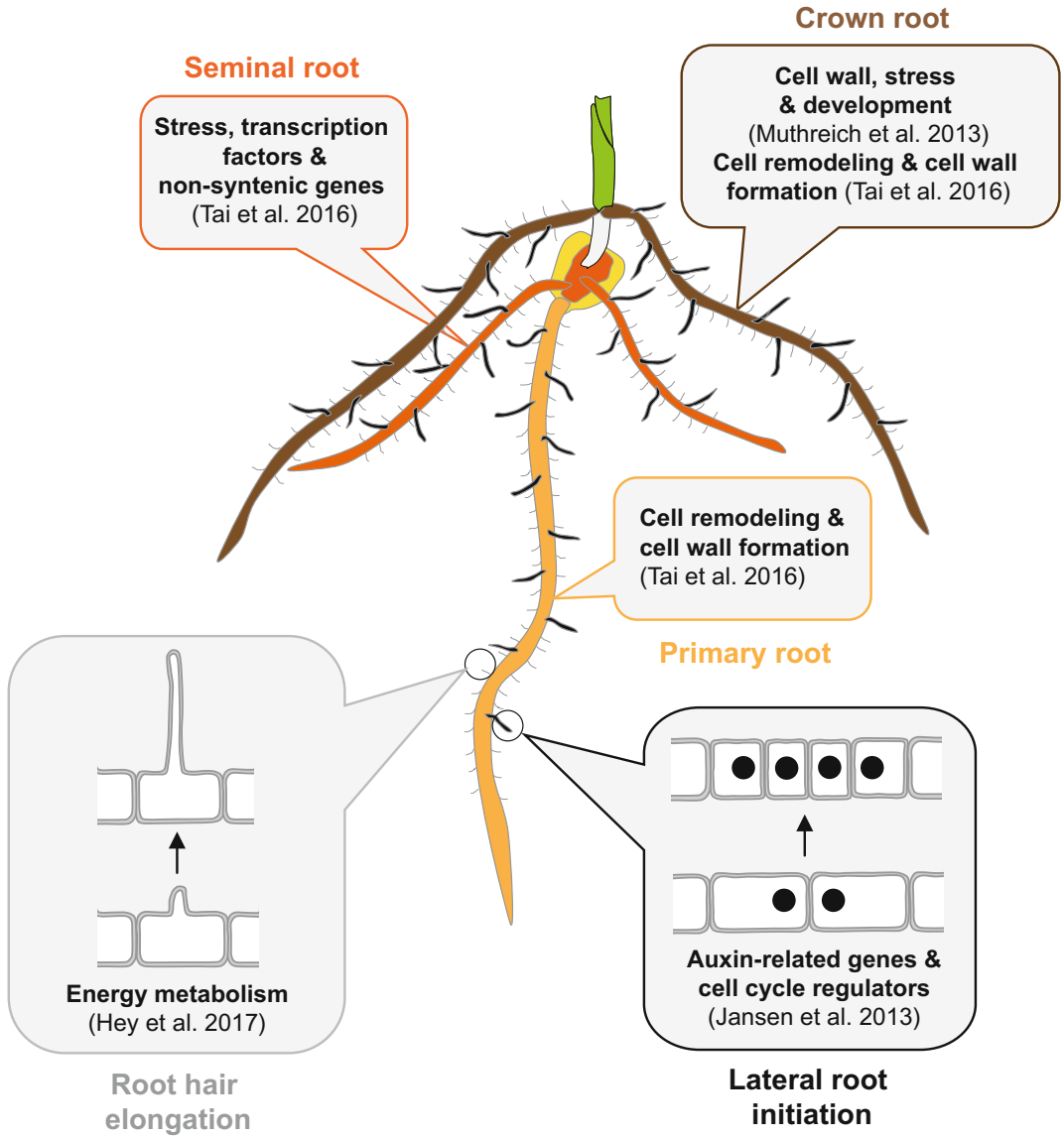

Fig. 15.1 Summary of transcriptome studies related to individual root types of maize

aboveground shoot nodes are brace roots. Shoot-borne roots take up most nutrients and are therefore critical for plant yield (Hochholdinger et al. 2004a). Genetic analyses demonstrated that root type-specific developmental programs control seminal, crown and lateral root formation in maize (Hochholdinger et al. 2004a).

The maize mutant *rtcs (rootless concerning crown and seminal roots)* does not initiate any shoot-borne roots (Hetz et al. 1996). A comparative microarray analysis of the first shoot-node of wild type and *rtcs* seedlings during three early stages of crown root initiation identified >800 transcripts that displayed RTCS-dependent expression. This suggests a direct or indirect role of these genes during shoot-borne root initiation (Muthreich et al. 2013). Functional classes of genes overrepresented among RTCS-dependent genes refer to cell wall, stress, and development-related processes (Muthreich et al. 2013).

A RNA-Seq study of primary, seminal and crown roots highlighted diverse transcriptomic patterns of these root types (Tai et al. 2016). Remarkably, embryonic primary and post-embryonic crown roots displayed similar anatomical features and transcriptomic profiles, while the transcriptome of seminal roots was distinct from these root types (Tai et al. 2016). In primary and crown roots, functions related to cell remodeling and cell wall formation were prominent. In contrast, in seminal roots stress-related genes and transcriptional regulators were overrepresented, which suggests a functional specialization of the different root types (Tai et al. 2016). Seminal roots also displayed unique anatomical features, which indicate a higher absorption efficiency in comparison with the other root types (Burton et al. 2013; Tai et al. 2016). In another RNA-Seq analysis of wild type and mutant *rtcs* embryos, which do not initiate seminal roots, it was demonstrated that evolutionary young non-syntenic genes were overrepresented among genes displaying RTCS-dependent expression during seminal root primordia formation (Tai et al. 2017). Non-syntenic genes evolved after the separation of the lineages leading to maize and sorghum (Schnable et al. 2011). It was suggested that these differentially expressed non-syntenic genes might have come under the transcriptional control of the syntenic gene *rtcs* during seminal root evolution (Tai et al. 2017).

A commonality of all embryonic and post-embryonic root types is their ability to initiate lateral roots and root hairs from pericycle and epidermis cells, respectively (Yu et al. 2016; Hochholdinger et al. 2018a). Continuous production of lateral roots and root hairs from the main root types substantially increases the root surface and thus facilitates the capturing of nutrient and water in maize (Rogers and Benfey 2015; Hochholdinger et al. 2018a). The maize mutant *lrt1 (lateralrootless 1)* is specifically affected in the formation of lateral roots on the embryonic primary root and seminal roots but not on the post-embryonic shoot-borne roots (Hochholdinger and Feix 1998). Similarly, the mutant *rum1 (rootless with undetectable meristem 1)* does not initiate lateral roots on the primary root but displays normal lateral root formation on crown and brace roots (Woll et al. 2005). These mutants demonstrate that root type-specific developmental programs control lateral root initiation in maize. The differential regulation of lateral root initiation in primary and crown roots of maize might be explained by members of root type-specific transcription factor families such as MYB (V-MYB AVIAN MYELOBLASTOSIS VIRAL ONCOGENE HOMOLOG), MYB-related, HOMEOBOX, and bHLH (basic helix-loop-helix) (Jansen et al. 2013).

15.2.2 Tissue Type-Specific Transcriptomics

Maize roots display highly differentiated tissue patterns along the longitudinal and radial axes. In longitudinal orientation, roots are structurally divided into the root cap at the terminal end, a subterminal meristematic zone, followed by the elongation and differentiation zone (Fig. 15.2; Ishikawa and Evans 1995; Hochholdinger et al. 2004a). These zones are functionally diversified

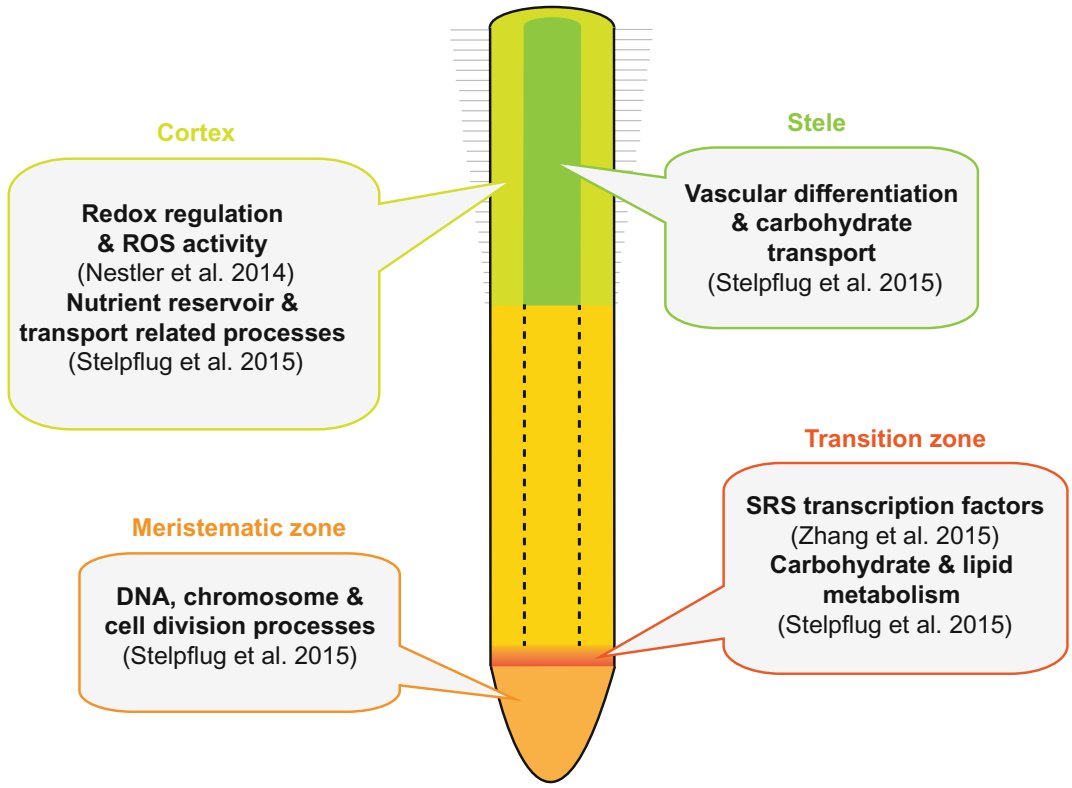

Fig. 15.2 Summary of transcriptome studies related to tissue patterning of the maize primary root. ROS, reactive oxygen species

by mitotic divisions in the meristematic zone, cell extension, and elongation in the elongation zone and highly differentiated cells with root hairs in the differentiation zone (Ishikawa and Evans 1995). In transverse orientation, the differentiation zone displays a number of functionally distinct cell types (Hochholdinger 2009; Yu et al. 2016). The stele, in the center of the root, contains differentiated xylem vessels and primary phloem elements. Xylem vessels are involved in water and nutrient transport, while primary phloem elements transport photosynthates. Both, xylem and phloem elements are embedded in the ground tissue of the central cylinder. The pericycle represents the outermost cell layer of the central cylinder. Phloem pole pericycle cells in the differentiation zone can divide and give rise to lateral roots. The pericycle is connected via a single layer of endodermis cells to the multilayered parenchyma. The outermost cell layer of the root is the epidermis that connects the root to the rhizosphere. The stele and the cell layers between endodermis and epidermis, hereafter referred to as cortical parenchyma, can be mechanically separated at the boundary between pericycle and endodermis without damaging these cell files, thus allowing subsequent analyses of these functionally diverse tissues (Saleem et al. 2009; Paschold et al. 2014; Yu et al. 2015). The functionally diverse cell types described above can be distinguished based on their position within the three-dimensional context of the root but also according to their anatomical characteristics and their staining ability by certain histological dyes (Hochholdinger 2009).

Maize root tissues are highly diversified in their biological functions along the longitudinal and transverse axes. A snapshot of the dynamics of maize root tissues was generated by co-expression analyses of 17 transcriptome

profiles generated from embryonic and post-embryonic root tissues (Stelpflug et al. 2015). Genes encoding enzymes associated with "translation", "ribosomal function and assembly", "protein metabolism", "DNA synthesis and replication", "transcriptional activation", "cell cycle regulation", "microtubule motor activity", "nucleosome assembly", and "cell wall organization" were highly enriched in the root tip. This is consistent with functions required by the rapidly dividing cells in the meristematic zone. Genes showing peak expression in the transition zone of the root, which is located between the meristematic and the elongation zone (Fig. 15.2), were enriched in "carbohydrate and lipid metabolic processes", "response to oxidative stress", "peroxidases", "lignin catabolism", and "cell wall organization" (Stelpflug et al. 2015). Moreover, SRS (SHI RELATED SEQUENCE) transcription factors , involved in auxin-mediated lateral root primordia initiation in maize were overrepresented in this region (Zhang et al. 2014). The differentiation zone featured predominant expression of genes encoding nutrient reservoir activity, transport, kinases, protein phosphorylation, regulation of transcription, and transcription factor activity (including enrichment for TIFY (TIFY DOMAIN), MYB, NAC (NAM, ATAF, and CUC) and WRKY (WRKY DOMAIN) families, monooxygenase activity, glutathione transferases, redox regulation, electron carrier activity, lipid metabolism, and biosynthesis of flavonoids (Stelpflug et al. 2015). Redox regulation and ROS (reactive oxygen species) activity are crucial for root hair initiation and elongation in maize, which occur in this zone (Nestler et al. 2014). In the cortical parenchyma, 61 GO (gene ontology) categories were enriched highlighting the prevalent biological functions in this complex tissue. In contrast, only six GO terms were enriched in the stele indicating the specialization of this tissue to processes such as vascular differentiation and carbohydrate transport (Stelpflug et al. 2015). Overall, GO enrichment analyses showed disparate patterns of gene expression between cortical parenchyma and stele, which is consistent with distinct biological functions of these tissues.

15.2.3 Cell Type-Specific Transcriptomics

Root hairs are tubular extensions of epidermis cells. Their easy accessibility made them a model for single cell-type research in plants. In maize, transcriptome profiling demonstrated that the root hair transcriptome is less complex than the transcriptome of primary roots without root hairs, which consists of multiple cell types (Hey et al. 2017). A substantial number of GO terms enriched in the transcriptome of root hairs are functionally related to "energy metabolism", which highlights the high-energy demand for the development of these cells and the maintenance of their function (Hey et al. 2017). LCM (laser capture microdissection) provides the opportunity to isolate specific cell types from inside complex tissues (Nakazono et al. 2003). RNA isolated from individual cells can be linearly amplified and subsequently subjected to downstream transcriptome analyses (Ludwig and Hochholdinger 2014). Early LCM studies in combination with microarray hybridization experiments revealed genes involved in the specification of pericycle cells (Dembinsky et al. 2007) and genes regulating the stages shortly before lateral root initiation (Woll et al. 2005). More recently, LCM-based pericycle-specific RNA-Seq analyses indicated that auxin-related genes and cell cycle regulators play a crucial role during lateral root initiation in maize (Yu et al. 2015). Notably, pericycle cells isolated from brace roots showed a unique transcriptomic landscape compared to other seedling root types, which is consistent with the distinct lateral root density in these root types (Yu et al. 2016).

15.3 Transcriptomic Responses to Different Supply of Mineral Nutrients in Maize Roots

The architectural modulation of the root system of modern maize in comparison to its teosinte counterpart can be explained in part by its adaptation to different local environmental conditions during domestication. Root morphology

Table 15.1 Summary of transcriptome studies related to heterosis, drought, hydropatterning, and nitrate. LBD, lateral organ boundary domain; NO, Nitric oxide; ROS, reactive oxygen species; SPE, single parent expression; PIN, pin-formed

Studied root type or zone	Factors	Enrichment of	References
Primary root	Heterosis	SPE and non-syntenic genes	Paschold et al. (2012, 2014), Baldauf et al. (2016, 2018), Marcon et al. (2017)
	Drought	Transcriptional regulation, ROS, and hormone metabolism	Opitz et al. (2014, 2016), Marcon et al. (2017)
Lateral root	Hydropatterning	Efflux and biosynthesis of auxin, LBD	Bao et al. (2014), Robbins and Dinneny (2018)
	Local nitrate	Auxin transport and PIN9	Liu et al. (2010), Yu et al. (2015)
Meristematic zone	Drought	Cell wall reorganization	Opitz et al. (2016)
Transition zone	Nitrate	Molecular NO signal	Trevisan et al. (2011), Manoli et al. (2014)

and architecture determine the potential to exploit unevenly distributed nutrient patches in the soil to maximize nutrient-use efficiency and yield (Yu et al. 2014). Physiological experiments suggest that maize plants optimize their root architecture by regulating lateral root formation based on the availability of soil nitrate. Sparsely spaced and long lateral roots are optimal for nitrate acquisition in maize (Lynch 2011, 2013). Recent transcriptome studies highlighted diverse transcriptomic changes in response to nitrogen-deficiency and nitrogen-enrichment conditions (Table 15.1).

15.3.1 Nitrate Deficiency Triggered Transcriptome Changes of Maize Root

Systemic nitrate deficiency in maize seedlings increases carbon partitioning to roots, accelerates root growth, and results in fewer, longer main roots with longer lateral roots (Gaudin et al. 2011; Trachsel et al. 2013). Recent RNA-Seq analyses revealed that several AP2-EREBP (APETALA2/ETHYLENE-RESPONSIVE ELEMENT BINDING PROTEIN) family members play important roles in the *rtcs*-dependent regulatory network related to both root development and nitrate-deficiency response in maize (He et al. 2015). Moreover, complex transcriptomic reprogramming occurs as an early response to nitrate supply and molecular NO (nitric oxide) in the transition zone between meristematic and elongation zone (Trevisan et al. 2011; Manoli et al. 2014). These results emphasize the role of the transition zone of maize roots in sensing and transducing nitrate signals (Trevisan et al. 2011, 2015).

15.3.2 Local Nitrate Enrichment Triggered Transcriptome Changes of Maize Roots

In resource-depleted environments, heterogeneously distributed nutrient availability directs lateral root growth preferentially into nutrient-rich patches (Forde 2014; Giehl and von Wirén 2014). The application of auxin and auxin transport inhibitors revealed the pivotal role of auxin shoot-to-root transport in lateral root growth in response to localized nitrate supply (Wang et al. 2005; Tian et al. 2008; Guo et al. 2005; Liu et al. 2010). Microarray analyses indicated that early responsive genes related to cell division and expansion, such as α-expansin, cellulose synthase, kinesin, plasma membrane and tonoplast aquaporins, are possibly involved in localized nitrate stimulation of lateral root development in maize (Wang et al. 2005; Liu

et al. 2010). Among the different root types of maize, brace roots displayed an exceptional plasticity for lateral root formation compared to other root types upon different nitrate concentrations. This is illustrated by an increased induction of lateral roots in brace roots by nitrate compared to other root types and is reflected by their unique transcriptional control of cell cycle and MYB-related genes (Yu et al. 2014, 2015, 2016). Notably, activity of the monocot-specific PIN9 (*PIN-FORMED 9*) gene in phloem pole cells modulates auxin efflux to pericycle cells and subsequent cell cycle activation (Yu et al. 2015). Taken together, these recent discoveries link transcriptional regulation and nutrient-triggered lateral root development in maize.

15.4 Transcriptomic Responses to Drought Stress in Maize Roots

In maize, a deeper root system enhances drought tolerance (Ribaut et al. 2009), while shallower seminal and crown roots increase phosphorus acquisition (Zhu et al. 2005). Recent studies indicated that seminal and crown roots are major targets for drought stress signaling in maize (Gao and Lynch 2016; Sebastian et al. 2016). To study the transcriptomic response to intensity and duration of drought stress simulated by PEG (polyethylene glycol) 8000 treatment, maize primary roots were subjected to mild (−0.2 MPa) and severe (−0.8 MPa) water deficit conditions for 6 h and 24 h (Opitz et al. 2014). In general, the number of drought-responsive genes increased with intensity and duration of water deficit. In this study, a set of 53 genes were drought-responsive independent of the type of water deficit treatment. Among the differentially expressed genes, the overrepresented GO categories "oxidoreductase activity" and "heme binding" connected water deficit response to ROS metabolism (Opitz et al. 2014). This study provided a first global insight into water deficit-responsive genes in young maize primary roots and provided candidate genes as a starting point for future genetic analyses. In a follow-up study, the transcriptomic dynamics of distinct tissues of maize primary roots in response to drought were surveyed by exposing seedlings to PEG 8000 treatment simulating a low water potential of −0.8 MPa for 6 h (Opitz et al. 2016). Comparison of gene expression between control and drought conditions in the meristematic zone, elongation zone and in cortex and stele of the differentiation zone revealed an remarkable transcriptomic plasticity of water deficit response in these tissues. The highest number of water deficit-responsive genes was detected in the cortex and the elongation zone. Furthermore, GO terms "transcriptional regulation" and "hormone metabolism" were most prominent among the differentially expressed genes in all tissues. This indicates a global reprogramming of cellular metabolism in adaptation to drought. Moreover, genes associated with cell wall reorganization were most abundant among differentially expressed genes in the root tip. These genes likely allow for continued root growth under water deficit conditions (Opitz et al. 2016).

It has been demonstrated that lateral root formation is repressed by transient water deficit in maize (Babé et al. 2012). Upon water deficit, pericycle founder cells are blocked irreversibly at the asymmetric division stage and at various later stages (Babé et al. 2012). Recently, it has been shown that roots of Arabidopsis, rice, and maize can sense heterogeneity in water availability by "hydropatterning" in transverse orientation at the sub-organ level (Bao et al. 2014). Hydropatterning is preceded by PIN-mediated auxin efflux and TAA1 (TRYPTOPHANE-PYRUVATE AMINOTRANSFERASE 1)-mediated auxin biosynthesis to determine pre-branch sites of lateral roots (Bao et al. 2014). The process of hydropatterning is independent of endogenous abscisic acid signaling, thus distinguishing it from a classical drought response (Bao et al. 2014). Transcriptomic characterization of the local response of maize lateral root formation in contact with a moist surface revealed extensive regulation of signaling pathways, such as upregulated genes encoding LOB (LATERAL ORGAN BOUNDARIES) domain transcriptional factors, including RTCS (Taramino et al. 2007; Robbins and Dinneny 2018).

15.5 Transcriptomic Plasticity During Heterosis Manifestation in Maize Hybrids

Maize hybrids perform significantly better than their parental inbred lines, a phenomenon known as heterosis (reviewed in Hochholdinger and Hoecker 2007; Schnable and Springer 2013). Young maize roots are an ideal system to study the manifestation of heterosis, because developmental differences between inbred lines and their hybrids are established within a few days after germination thus allowing these plants to grow under very controlled conditions (Hoecker et al. 2006). An early transcriptome study based on 12 k maize microarray chips demonstrated genotype- and organ-specific non-additive gene expression patterns in hybrids (Hoecker et al. 2008). More recently, RNA-Seq-based transcriptome analyses highlighted the transcriptomic plasticity of global gene expression patterns in primary roots of maize inbred lines versus their hybrids. It was demonstrated that hundreds of genes display an extreme instance of gene expression complementation by being expressed in both hybrids but in only one of the parental inbred lines (Paschold et al. 2012). These expression patterns were designated SPE (single parent expression) complementation. Such complementation patterns are consistent with the dominance model of heterosis (Jones 1917). In subsequent studies, it was demonstrated that SPE complementation patterns are highly dynamic in different root tissues (Paschold et al. 2014) and stable under water deficit stress (Marcon et al. 2017). The concept of SPE complementation was recently generalized by demonstrating that a panel of diverse inbred lines and their hybrids consistently displayed hundreds of SPE patterns during three stages of primary root development (Baldauf et al. 2018). Remarkably, evolutionary younger non-syntenic genes were significantly overrepresented among genes displaying SPE patterns (Paschold et al. 2014; Baldauf et al. 2018). It was also demonstrated that non-syntenic genes drive non-additive, allele-specific, and differential gene expression patterns in distinct tissues of maize primary roots (Baldauf et al. 2016). Furthermore, it was indicated that non-syntenic genes contribute to the adaption of plants under changing environmental conditions (Marcon et al. 2017). A possible function of non-syntenic SPE genes could therefore be associated with the increased potential of adaptation of hybrids to environmental changes.

15.6 Future Prospects

Functional genomics studies and systems biology approaches increasingly contribute to the understanding of plant biology. Integration of functional genomics and high-throughput phenotyping will help to better understand the molecular processes during maize root development. Currently, most transcriptome analyses are executed at the level of entire plants, organs, or tissues. Although these experiments contribute to the understanding of the transcriptomic landscapes of plants, they do not resolve the cellular complexity of the analyzed samples. Techniques to isolate and subsequently analyze the biomolecules of single-cell types provide unparalleled opportunities for high-resolution systemic dissection of root traits and complex biological phenomena such as heterosis. Finally, plant roots intensively interact with their environment via the rhizosphere, which is the narrow proportion of soil that is influenced by root secretions and soil microorganisms. A grand challenge of root biology is to decipher the complex interplay between root systems and microbes in the rhizosphere of natural habitats.

Acknowledgements Root research in FH's laboratory supported by the DFG (German Research Foundation) and the BMBF (German Federal Ministry of Education and Research).

Author Contributions All authors contributed to the writing of this book chapter.

References

Babé A, Lavigne T, Severin J-P, Nagel KA, Walter A, Chaumont F, Batoko H, Beeckman T, Draye X (2012) Repression of early lateral root initiation events by transient water deficit in barley and maize. Philos Trans R Soc B Biol Sci 367:1534–1541

Baldauf JA, Marcon C, Lithio A, Vedder L, Altrogge L, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F (2018) Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. Curr Biol 28:431–437

Baldauf JA, Marcon C, Paschold A, Hochholdinger F (2016) Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive and allelic expression patterns in maize hybrids. Plant Physiol 171:1144–1155

Bao Y, Aggarwal P, Robbins NE, Sturrock CJ, Thompson MC, Tan HQ, Tham C, Duan L, Rodriguez PL, Vernoux T et al (2014) Plant roots use a patterning mechanism to position lateral root branches toward available water. Proc Natl Acad Sci 111:9319–9324

Burton AL, Brown KM, Lynch JP (2013) Phenotypic diversity of root anatomical and architectural traits in *Zea* species. Crop Sci 53:1042–1055

Dembinsky D, Woll K, Saleem M, Liu Y, Fu Y, Borsuk LA, Lamkemeyer T, Fladerer C, Madlung J, Barbazuk B et al (2007) Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. Plant Physiol 145: 575–588

Forde BG (2014) Nitrogen signalling pathways shaping root system architecture: an update. Curr Opin Plant Biol 21:30–36

Gao Y, Lynch JP (2016) Reduced crown root number improves water acquisition under water deficit stress in maize (*Zea mays* L.). J Exp Bot 67:4545–4557

Gaudin ACM, Mcclymont SA, Holmes BM, Lyons E, Raizada MN (2011) Novel temporal, fine-scale and growth variation phenotypes in roots of adult-stage maize (*Zea mays* L.) in response to low nitrogen stress. Plant Cell Environ 34:2122–2137

Giehl RFH, von Wiren N (2014) Root nutrient foraging. Plant Physiol 166:509–517

Guo Y, Chen F, Zhang F, Mi G (2005) Auxin transport from shoot to root is involved in the response of lateral root growth to localized supply of nitrate in maize. Plant Sci 169:894–900

He X, Qu B, Li W, Zhao X, Teng W, Ma W, Ren Y, Li B, Li Z, Tong Y (2015) The nitrate-inducible NAC transcription factor TaNAC2-5A controls nitrate response and increases wheat yield. Plant Physiol 169:1991–2005

Hetz W, Hochholdinger F, Schwall M, Feix G (1996) Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. Plant J 10:845–857

Hey S, Baldauf JA, Opitz N, Lithio A, Pasha A, Provart N, Nettleton D, Hochholdinger F (2017) Complexity and specificity of the maize (*Zea mays* L.) root hair transcriptome. J Exp Bot 68:2175–2185

Hochholdinger F (2009) The maize root system: morphology, anatomy, and genetics. In: Bennetzen JL, Hake SC (eds) Handb. Maize, Springer, pp 145–160

Hochholdinger F, Feix G (1998) Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. Plant J 16:247–255

Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. Trends Plant Sci 12:427–432

Hochholdinger F, Marcon C, Baldauf JA, Frey F, Yu P (2018b) Proteomics of maize root development. Front Plant Sci 9:143

Hochholdinger F, Park WJ, Sauer M, Woll K (2004a) From weeds to crops: genetic analysis of root development in cereals. Trends Plant Sci 9:42–48

Hochholdinger F, Woll K, Sauer M, Dembinsky D (2004b) Genetic dissection of root formation in maize (*Zea mays*) reveals root type specific developmental programmes. Ann Bot 93:359–368

Hochholdinger F, Yu P, Marcon C (2018a) Genetic control of root system development in maize. Trends Plant Sci 23:79–88

Hoecker N, Keller B, Muthreich N (2008) Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics 179:1275–1283

Hoecker N, Keller B, Piepho HP, Hochholdinger F (2006) Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theor Appl Genet 112:421–429

Ishikawa H, Evans ML (1995) Specialized zones of development in roots. Plant Physiol 109:725–727

Jansen L, Hollunder J, Roberts I, Forestan C, Fonteyne P, Van Quickenborne C, Zhen RG, Mckersie B, Parizot B, Beeckman T (2013) Comparative transcriptomics as a tool for the identification of root branching genes in maize. Plant Biotechnol J 11:1092–1102

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C et al (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527

Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. Genetics 2:466–479

Liu J, An X, Cheng L, Chen F, Bao J, Yuan L, Zhang F, Mi G (2010) Auxin transport in maize roots in response to localized nitrate supply. Ann Bot 106:1019–1026

Ludwig Y, Hochholdinger F (2014) Laser microdissection of plant cells. In: Žárský V, Cvrčková F (eds) Plant cell morphogenesis. Methods in molecular biology. Humana Press, Totowa, pp 249–258

Lynch JP (2011) Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. Plant Physiol 156:1041–1049

Lynch JP (2013) Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. Ann Bot 112:347–357

Manoli A, Begheldo M, Genre A, Lanfranco L, Trevisan S, Quaggiotti S (2014) NO homeostasis is a key regulator of early nitrate perception and root elongation in maize. J Exp Bot 65:185–200

Marcon C, Paschold A, Malik WA, Lithio A, Baldauf JA, Altrogge L, Opitz N, Lanz C, Schoof H, Nettleton D et al (2017) Stability of single parent gene expression complementation in maize hybrids upon water deficit stress. Plant Physiol 173:1247–1257

Muthreich N, Majer C, Beatty M, Paschold A, Schutzenmeister A, Fu Y, Malik WA, Schnable PS, Piepho H-P, Sakai H et al (2013) Comparative transcriptome profiling of maize coleoptilar nodes during shoot-borne root initiation. Plant Physiol 163:419–430

Nakazono M, Qiu F, Borsuk LA, Schnable PS (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 15:583–596

Nestler J, Liu S, Wen TJ, Paschold A, Marcon C, Tang HM, Li D, Li L, Meeley RB, Sakai H et al (2014) *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. Plant J 79:729–740

Opitz N, Marcon C, Paschold A, Malik WA, Lithio A, Brandt R, Piepho HP, Nettleton D, Hochholdinger F (2016) Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. J Exp Bot 67:1095–1107

Opitz N, Paschold A, Marcon C, Malik WA, Lanz C, Piepho H-P, Hochholdinger F (2014) Transcriptomic complexity in young maize primary roots in response to low water potentials. BMC Genom 15:741

Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh C-T, Ossowski S, Lanz C, Nettleton D, Schnable PS et al (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res 22:2445–2454

Paschold A, Larson NB, Marcon C, Schnable JC, Yeh CT, Lanz C, Nettleton D, Piepho H-P, Schnable PS, Hochholdinger F (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell 26:3939–3948

Peter R, Eschholz TW, Stamp P, Liedgens M (2009) Early growth of flint maize landraces under cool conditions. Crop Sci 49:169–178

Ribaut JM, Betran J, Monneveux P, Setter T (2009) Drought tolerance in maize. In: Bennetzen JL, Hake SC (eds) Handb. Maize, Springer, pp 311–344

Robbins NE, Dinneny JR (2018) Growth is required for perception of water availability to pattern root branches in plants. Proc Natl Acad Sci 201710709

Rogers ED, Benfey PN (2015) Regulation of plant root system architecture: implications for crop advancement. Curr Opin Biotechnol 32:93–98

Saleem M, Lamkemeyer T, Schützenmeister A, Fladerer C, Piepho HP, Nordheim A, Hochholdinger F (2009) Tissue specific control of the maize (*Zea mays* L.) embryo, cortical parenchyma, and stele proteomes by RUM1 which regulates seminal and lateral root initiation. J Proteome Res 8:2285–2297

Schnable PS, Springer NM (2013) Progress toward understanding heterosis in crop plants. Annu Rev Plant Biol 64:13.1–13.18

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Nat Acad Sci 108:4069–4074

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Sebastian J, Yee M-C, Goudinho Viana W, Rellán-Álvarez R, Feldman M, Priest HD, Trontin C, Lee T, Jiang H, Baxter I et al (2016) Grasses suppress shoot-borne roots to conserve water during drought. Proc Natl Acad Sci 113:8861–8866

Stelpflug SC, Rajandeep S, Vaillancourt B, Hirsch CN, Buell CR, De Leon N, Kaeppler SM (2015) An expanded maize gene expression atlas based on RNA-sequencing and its use to explore root development. Plant Genome 9:314–362

Tai H, Lu X, Opitz N, Marcon C, Paschold A, Lithio A, Nettleton D, Hochholdinger F (2016) Transcriptomic and anatomical complexity of primary, seminal, and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.). J Exp Bot 67:1123–1135

Tai H, Opitz N, Lithio A, Lu X, Nettleton D, Hochholdinger F (2017) Non-syntenic genes drive RTCS-dependent regulation of the embryo transcriptome during formation of seminal root primordia in maize (*Zea mays* L.). J Exp Bot 68:403–414

Taramino G, Sauer M, Stauffer JL, Multani D, Niu X, Sakai H, Hochholdinger F (2007) The maize (*Zea mays* L.) RTCS gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. Plant J 50:649–659

Tian Q, Chen F, Liu J, Zhang F, Mi G (2008) Inhibition of maize root growth by high nitrate supply is correlated with reduced IAA levels in roots. J Plant Physiol 165:942–951

Trachsel S, Kaeppler SM, Brown KM, Lynch JP (2013) Maize root growth angles become steeper under low N conditions. Field Crop Res 140:18–31

Trevisan S, Manoli A, Begheldo M, Nonis A, Enna M, Vaccaro S, Caporale G, Ruperti B, Quaggiotti S (2011) Transcriptome analysis reveals coordinated spatiotemporal regulation of hemoglobin and nitrate reductase in response to nitrate in maize roots. New Phytol 192:338–352

Trevisan S, Manoli A, Ravazzolo L, Botton A, Pivato M, Masi A, Quaggiotti S (2015) Nitrate sensing by the maize root apex transition zone: a merged transcriptomic and proteomic survey. J Exp Bot 66:3699–3715

Wang Y, Mi G, Chen F, Zhang J, Zhang F (2005) Response of root morphology to nitrate supply and its

contribution to nitrogen accumulation in maize. J Plant Nutr 27:2189–2202

Woll K, Borsuk L a, Stransky H, Nettleton D, Schnable PS, Hochholdinger F (2005) Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. Plant Physiol 139:1255–1267

Yu P, Baldauf JA, Lithio A, Marcon C, Nettleton D, Li C, Hochholdinger F (2016) Root type-specific reprogramming of maize pericycle transcriptomes by local high nitrate results in disparate lateral root branching patterns. Plant Physiol 170:1783–1798

Yu P, Eggert K, von Wirén N, Li C, Hochholdinger F (2015) Cell type-specific gene expression analyses by RNA sequencing reveal local high nitrate-triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle regulation. Plant Physiol 169:690–704

Yu P, White PJ, Hochholdinger F, Li C (2014) Phenotypic plasticity of the maize root system in response to heterogeneous nitrogen availability. Planta 240:667–678

Zhang Y, Paschold A, Marcon C, Liu S, Tai H, Nestler J, Yeh C, Opitz N, Lanz C, Schnable PS et al (2014) The *Aux/IAA* gene rum1 involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. J Exp Bot 65:4919–4930

Zhu J, Kaeppler SM, Lynch JP (2005) Topsoil foraging and phosphorus acquisition efficiency in maize (*Zea mays*). Funct Plant Biol 32:749–762

Genomics of Nitrogen Use Efficiency in Maize: From Basic Approaches to Agronomic Applications

16

Bertrand Hirel and Peter J. Lea

Abstract

Maize farming requires high amounts of nitrogen (N) fertilizer, which can have detrimental effects on agronomic sustainability and the environment. Thus, irrespective of the mode of N fertilization, an increased knowledge of the mechanisms controlling plant N metabolism is essential for improving nitrogen use efficiency (NUE) in maize. This new knowledge will reduce the excessive input of fertilizers, while maintaining an acceptable yield and a sufficient profit margin for the farmers. It is now possible to further develop whole-plant agronomic and physiological studies. These can be combined with gene, protein, and metabolite profiling to build up a comprehensive picture depicting the different steps of N uptake, assimilation, and recycling to produce either biomass in vegetative organs or proteins in storage organs. We provide an overview describing how our understanding of the physiological and molecular controls of N assimilation in maize has been advanced using combined approaches. These are based on agronomic, whole-plant physiology, genetic, modeling, and systems biology approaches. Current knowledge and prospects for selecting high-yielding maize genotypes adapted to lower N fertilizer input and for identifying biological markers representative of the plant N status for breeding and agronomic purposes are reviewed.

16.1 Introduction: Nitrogen Fertilization and Sustainable Food Production

Today, the application of mineral fertilizers such as nitrogen (N) is the main agricultural practice used to maintain and restore soil nutrients and thus stabilize or even increase crop yields. In commercial fertilizers, the applied N is particularly soluble for easy uptake by plants allowing its rapid assimilation during root and shoot vegetative growth. The storage and handling of mineral N fertilizers are relatively easy, thus allowing application before crop growth or using fractionated applications in certain species notably when the plant needs it most (Reetz et al. 2015). Mineral fertilizers are now the main source of nutrients applied to soils, although animal manures are also commonly used to

B. Hirel (✉)
Adaptation des Plantes à Leur Environnement, Unité Mixte de Recherche 1318, Institut Jean-Pierre Bourgin, Institut National de La Recherche Agronomique, Centre de Versailles-Grignon, R.D. 10, 78026 Versailles Cedex, France
e-mail: bertrand.hirel@inra.fr

P. J. Lea
Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_16

fertilize crops. After the Second World War, instead of enriching the soil stock with organic C and N from farmyard manures, leguminous cultures, fallow periods, and synthetic N fertilizers have been extensively used to increase crop yield. The chemical synthesis of ammonium for the production of mineral N fertilizers by the Haber–Bosch process allowed increasing agricultural production to feed the constantly growing world population (Galloway et al. 2013). Consequently, it has been estimated that since 1961, the total N applied to crops has increased dramatically by a factor of 4.4. However, total protein production has only increased by a factor of 3.1, thus indicating that nitrogen use efficiency (NUE), defined as the yield obtained/unit of available N in the soil (supplied by the soil + N fertilizer), has declined substantially from 66 to 46% (Lassaletta et al. 2016). Plant NUE is the product of N uptake efficiency (amount of N taken up/quantity of available N) and of the N utilization efficiency (yield/absorbed N), (Pathak et al. 2011). There is a large genetic variability for both N uptake efficiency and N utilization efficiency in many crops, notably in maize (Han et al. 2015b). However, in many cases the best-performing maize varieties at high N fertilization input are not necessarily the best ones when the supply of N is reduced (Gallais and Coque 2005), depending on the world region, N management strategies and environmental factors including plant water status (Fixen et al. 2015). This is likely because most of the breeding strategies have been conducted under non-limiting N fertilization conditions, thus missing the opportunity of selecting for high-yielding maize varieties when the N fertilization conditions were low (Haegele et al. 2013). High N fertilization rates are used in most high-yielding intensive agricultural production systems. However, under such growth conditions over 50% and up to 75% of the mineral N applied to the field is not taken up by the plant and is lost by leaching into the soil (Cameron et al. 2013). Mineral N, especially nitrate (NO_3^-) and urea {$CO(NH_2)_2$}, contains very soluble compounds that can run off into the surface water or flow into the groundwater. It has been argued that at high concentrations in drinking water and plant food sources, nitrate can be a serious risk for human health (Espejo-Herrera et al. 2016). Although the impact of N consumption remains controversial (Horde and Conley 2017), fertilizer usage directives have been established in several countries (EEA 2012; Galloway et al. 2013). On top of toxicological implications for animals and humans (Habermeyer et al. 2015), the presence of nitrate in the soil has an important impact on the soil microbial environment (Verzeaux et al. 2016a) and on other living organisms, due mainly to the eutrophication of freshwater and marine ecosystems (Moss 2008; Withers et al. 2014). Moreover, in eutrophic soils there is emission of nitrous oxide (N_2O), which has a global warming potential almost 300 times that of CO_2 (Fowler et al. 2015; Oita et al. 2016; Smith 2017), thus increasing the greenhouse effect. In these eutrophic soils, the emission of toxic ammonia (NH_3) into the atmosphere can also contribute to a process of acidification (Cameron et al. 2013; Galloway et al. 2013; Fowler et al. 2015). A further key point is that the chemical synthesis of N fertilizers requires the use of a considerable amount of energy (Gellings and Parmenter 2016) and can for maize production represent up to 25% of the operational cost for the farmer (https://www.extension.iastate.edu/agdm/crops/pdf/a1-20.pdf). This is why improving N management practices is recommended for the benefit of both the farmer and the fertilizer industry, while at the same time limiting water and air pollution (Kanter et al. 2015). Therefore, both NUE and energy input for N fertilizer production are seen as important indicators for the environmental impact of the production of most conventional and energy crops (Oita et al. 2016).

In order to feed the nine billion people projected for the world population in 2050, it will be necessary to increase agricultural production by 70–100% (McKenzie and Williams 2015). This increase in production must be obtained mainly in developing countries, but also in other countries that use intensive agriculture. This implies that the use of N fertilizers must also be increased, while maintaining the equilibrium of

global ecosystems. Sustainable agricultural practices such as fertilizer use rationalization, crop rotation, establishment of ground cover, and burial of crop residues in association with no-till practices are promising ways of overcoming the detrimental impact of the overuse of N fertilizers (Hirel et al. 2011; Reetz et al. 2015). Moreover, the use of genotypes more adapted to such sustainable practices appears to be a promising strategy to improve crop NUE (Ceccarelli 2014; Swain et al. 2014).

Both genetic manipulation and breeding are two other strategies that have been used to produce crops that can take up more N from the soil and utilize the absorbed N more efficiently (Pathak et al. 2011; Haegele et al. 2013; Han et al. 2015b). Among these breeding strategies, the selection of varieties that develop more efficient symbioses with arbuscular mycorrhizal fungi (AMF) may be an interesting alternative for increasing plant NUE (Hodge and Storer 2015; Verzeaux et al. 2017). In line with this strategy, it has been shown that the use of no-till and continuous cover cropping significantly increases the diversity of plant colonization by AMF in comparison with conventional tillage (Verzeaux et al. 2016b, 2017). Thus, these alternative farming techniques could also be an attractive way to increase NUE for a number of crops including maize, through the beneficial impact of AMF (Habbib et al. 2016, 2017).

Another promising alternative to improve crop NUE is to exploit further the N_2-fixing ability of bacteria colonizing the roots of cereals (Parnell et al. 2016). These N_2-fixing (i.e., diazotrophic) bacteria, which have been studied for over a hundred years, do not elicit the formation of symbiotic root nodules as shown by rhizobia in legumes. They are able to colonize the root surface and often also inner root tissues (Richardson et al. 2009) and provide substantial amounts of N to the host plant (Chalk 2016; Kuan et al. 2016). Some of these bacteria are sold commercially as biofertilizers or phytostimulators, especially in Latin America using different types of inoculant carriers (Bashan et al. 2014; de Souza et al. 2015). Other plant growth-promoting bacteria (PGPB) release hormones that can stimulate root development, thus increasing mineral nutrient acquisition including N (Cassán and Diaz-Zorita 2016).

16.2 Nitrogen Management in Maize

For most crop species including maize, the plant developmental cycle can be roughly divided into two main phases. The first one corresponds to the vegetative growth phase, during which young developing roots and leaves efficiently absorb and assimilate inorganic N (nitrate and ammonium) for amino acid and protein synthesis (Andrews et al. 2013). Amino acids and proteins are further used to build up both plant structures and the cellular machinery. This allows the uptake and transport of minerals and the assimilation of the carbon (C) and N necessary for plant growth and development until plant maturity (Hirel et al. 2007b; Fageria 2008). The second phase corresponds to the remobilization period when senescing tissues export organic metabolites to ensure the formation of new developing storage organs (Hirel et al. 2007b). These storage organs are involved in plant survival and reproduction and are represented by the ear and kernels in maize. Kernels contain C and N reserves that are used during germination to allow seedling development (Limami et al. 2002), and can act as sources of carbohydrates and proteins for food, energy, and forage (Klopfenstein et al. 2013). In maize, depending on the genotype notably in hybrids (Coque and Gallais 2007; Ciampitti and Vyn 2013), 45–65% of the grain N is provided from preexisting N in the stover before silking. The remaining 35–55% of the grain N originates from post-silking N uptake (Gallais and Coque 2005). Under field conditions, only a single application of N fertilizer is generally performed at sowing, in order to obtain optimal yields depending both on the genotype and on the residual N in the soil. However, when considering pre- and post-silking uptake, N can be fractionated by applying the N fertilizer after sowing at different leaf developmental stages depending on the degree of N

availability. Deficiency in N can depend on both plant demand and soil N availability during the whole-plant developmental cycle (Binder et al. 2000; Plénet and Lemaire 1999).

Using maize both as a model plant and as a crop, numerous agronomic, physiological, and molecular genetic studies have been carried out in order to increase our knowledge of the regulatory mechanisms controlling N assimilation and recycling and their relationships during plant growth and development. Such studies were conducted not only to increase our fundamental knowledge of N assimilation in a C_4 plant but also to identify agronomic, physiological, and genetic markers which could then be used for breeding new genotypes exhibiting a better NUE (Hirel and Gallais 2011). Studies using maize as a model crop have taken advantage of its wide genetic diversity, the availability of mutant collections, recombinant inbred lines, and straightforward transformation protocols. In addition, physiological, biochemical, and "omics" data, genome sequences (Hirel et al. 2007a), and more recently, genome-scale metabolic models (Simons et al. 2014a, b), have become available.

16.3 Nitrogen Uptake and Assimilation in Maize and Their Potential Improvement

16.3.1 Nitrogen Uptake and Reduction

During the assimilation phase, nitrate is the preferred N source for most wild and crop species such as maize, whether inorganic or organic N is provided to the plant (Bloom 2015). Nitrate is taken up by means of specific transporters located in the root cell membrane, having both low and high affinities for the anion. In higher plants, these transporters are represented by two main gene families, namely the nitrate transporter/peptide transporter (*NRT1/PTR*) family (*NPF*), and the *NRT2* family also called the major facilitator superfamily (*MFS*; Léran et al. 2014).

The availability of the maize genome sequence has allowed the identification of several members of the nitrate transport system. This has improved our understanding of how the nitrate uptake system contributes to net N uptake in response to a reduced N supply during plant development. Although high- and low-affinity nitrate transporters were identified in maize roots, the nitrate uptake capacity of the plant over 40 days after leaf emergence was only correlated with the accumulation of transcripts of *ZmNRT2.1 and ZmNRT2.2* encoding two high-affinity transporters, which were the most abundant compared to the other members of the multigene family (Garnett et al. 2013). In further studies, it was suggested that ZmNRT2.2 plays a major role in different maize genotypes by maintaining biomass production at both low and high N supply (Garnett et al. 2015). Patents providing methods to alter nitrate transporter activity using genetically modified plants have been submitted. Maize plants transformed with nucleotide sequences encoding yeast (Wang and Loussaert 2015) or maize nitrate transporters (Allen et al. 2014) exhibited an increased yield.

Therefore, using nitrate transporters as targets for future breeding strategies aimed at improving plant N uptake and thus NUE in maize and other cereals (Fan et al. 2016) appears to be promising. However, further work is still required to identify if, as with Arabidopsis, there are mechanisms by which nitrate is transported and sensed and how nitrate sensing crosstalks with developmental and hormonal responses as functions of plant demand and N availability (O'Brien et al. 2016). Recently, it has been shown that in tobacco plants overexpressing the maize nitrate transporter ZmNrt2.1, the transgenic plants exhibited altered expression pattern of a set of genes that are modulated by nitrate and calcium. This finding suggests that the maize high-affinity nitrate transporter plays a regulatory role in the overall plant gene expression system involving both calcium and nitrate sensing (Ibrahim et al. 2017). Coumarin, a secondary metabolite secreted by plants, is known to have an effect on root morphology. This allelopathic compound, in

conjunction with nitrate, has an inducible effect on the maize high-affinity transport system, indicating that the regulation of root development in relation to nitrate uptake is worth to be further investigated (Lupini et al. 2018). In addition, optimizing nitrate uptake in the different maize root classes, notably seminal and lateral roots, is another possible breeding goal to increase nitrate acquisition (York et al. 2016). In primary roots, the apical section seems to be involved in nitrate sensing and signaling, whereas the basal section is implicated in nitrate acquisition. However, in other reports it has been found that it is the root tip that exhibits a higher capacity to absorb nitrate (Sorgonà et al. 2011), a capacity which probably depends on the age of the plant. This latter process is via the co-regulated expression of an essential functional partner (NAR2 protein) of the inducible high-affinity nitrate transport system NRT2.1 (Lupini et al. 2016), which will also need to be taken into account for optimizing N uptake. Very recently, it has been shown that the two maize NPF6 transporters, Zm-NPF6.4 and Zm-NPF6.6, are permeable to both nitrate and chloride, the latter having a variable substrate selectivity when roots are exposed to high nitrate. Such a finding suggests that changing the substrate selectivity of these transporters using targeted mutagenesis could be a way of improving salt tolerance in plants (Wen et al. 2017).

In addition, studies on the interaction between nitrate uptake and root architecture usually performed on young developing maize plants should be extended to the post-silking period, which is particularly important in maize, since half of the N translocated to the grain is taken up during grain filling. Such investigations have generally been limited by the fact that in maize the large and complex root system is not easily accessible under field growth conditions (Yu et al. 2014). However, the recent study of York et al. (2016) indicates that genetic dissection of the spatiotemporal variations of nitrate uptake in relation to root architecture can be developed using ultra-high-density mapping of maize populations (Song et al. 2016). Ultimately, such a genetic dissection will allow the identification of the genes or loci involved in the control of nitrate uptake in relation to root architecture. This will take into account the dominance of the parental lines concerning both the architecture of the roots and the preference for nitrate or ammonium, in particular when hybrids are produced (Dechorgnat et al. 2018).

Ammonium (NH_4^+) is the ultimate form of inorganic N available to the plant. Most of the ammonium incorporated by the plant into organic molecules originates from nitrate reduction. However, metabolic pathways such as photorespiration, phenylpropanoid metabolism, utilization of N transport compounds, and amino acid catabolism can also generate considerable amounts of ammonium (Lea and Miflin 2011; Andrews et al. 2013).

In a cultivated soil, the ammonium concentration is generally much lower than that of nitrate (Nieder et al. 2011), but substantial amounts of ammonium can remain in the soil and may be used by crops despite active nitrification by soil microorganisms, which compete with plants for the acquisition of N (Bloom 2015). In higher plants, ammonium is taken up via an active transport system by means of proteins belonging to the ammonium transporter/methylammonium permease/Rhesus (AMT/MEP/Rh) family, located in the root cell plasma membrane and exhibiting different cellular distribution and substrate affinities (Yuan et al. 2007). Despite the importance of ammonium as a N source, it is only very recently that the characterization of the maize AMT-mediated high-affinity ammonium uptake mechanism has been carried out. Two ammonium transporters, ZmAMT1;1a and ZmAMT1;3, localized in the rhizodermis were identified as being the major components of the high-affinity transport system in maize roots. The two genes encoding the transporters are persistently induced by ammonium rather than up-regulated under ammonium deficiency, thus allowing the capture of low concentrations of the ammonium ion from the soil, even in the presence of nitrate (Gu et al. 2013). However, investigations into the level of expression, of the genes encoding ZmAMT1;1a and ZmAMT1;3, have failed to determine to what extent the two transporters contribute to the

overall N uptake and NUE of maize. The same question also applies to the role of urea transporters, which could play a role in the uptake and use of N in maize, notably under N-deficiency conditions (Liu et al. 2015a), as claimed in a patent in which the genetic manipulation of plants for efficient uptake and utilization of urea was described (Gupta et al. 2013).

Following uptake, nitrate is reduced first by the enzyme nitrate reductase (NR; E.C. 1.6.6.1) which catalyzes the reduction of nitrate to nitrite (NO_2^-) in the cytosol. Nitrate can be assimilated directly in the roots but is generally transported to the shoots for reduction in the leaf mesophyll cells in most herbaceous species including maize (Mengel et al. 1983). Subsequently, the enzyme nitrite reductase (NiR; E.C. 1.7.7.1) catalyzes the reduction of nitrite to ammonium in the plastids of these cells (Becker et al. 1993). Although several attempts have been made to manipulate the expression of the two enzymes involved in nitrate and nitrite reduction in a number of plant species (Pathak et al. 2011; Davenport et al. 2015), there is no published phenotypic characterization of transgenic maize plants in which a plant NR or NiR was overexpressed. There is a considerable amount of physiological and genetic evidence that levels of NR activity do not limit yield (Hirel et al. 2001; Andrews et al. 2004). However, it was recently reported that overexpression of a NR gene isolated from the alga *Porphyra perforata* (from which the gene sequence has been optimized for maize), simultaneously with a nitrate transporter from yeast, increased the yield of maize (Wang and Lous-saert 2015).

16.3.2 Ammonium Assimilation and Amino Acid Metabolism

In addition to nitrate reduction, ammonium can be generated by a variety of metabolic pathways such as photorespiration, phenylpropanoid metabolism, utilization of nitrogen transport compounds, and amino acid catabolism (Hirel and Lea 2002; Lea and Miflin 2011; Andrews et al. 2013). However, these metabolic processes occurring inside the plant may be qualitatively or quantitatively different in a C_4 plant such as maize compared to a C_3 plant (Oaks 1994; Bräutigam and Gowik 2016).

Irrespective of their origin, ammonium ions are then incorporated into the amide position of the amino acid glutamine in the presence of glutamate by the enzyme glutamine synthetase (GS; E.C. 6.3.1.2). The reaction catalyzed by the enzyme GS is the major, if not the only route facilitating the incorporation of inorganic N into organic molecules in conjunction with a second enzyme ferredoxin-dependent glutamate synthase (Fd-GOGAT; E.C. 1.4.7.1). Fd-GOGAT recycles glutamate and incorporates carbon skeletons into the GS/GOGAT cycle using the organic acid 2-oxoglutarate as a substrate (Fig. 16.1). Both glutamine and glutamate are later used as amino group donors for the biosynthesis of most of the other amino acids. These amino acids can be directly used for the synthesis of proteins and nucleotides, the latter being basic components of nucleic acids (Hirel and Lea 2001; Forde and Lea 2007) or further transported through the phloem stream, thus providing organic N to developing organs (Yesbergenova-Cuny et al. 2016). In C_3 and C_4 plants, both GS and GOGAT are present as several isoenzymes located in different cellular compartments and different cell types, according to the developmental stage of the plant. Typically, GS exists as two major isoforms: a cytosolic form occurring in the cytoplasm (GS1) and a plastidic form (GS2) present in the chloroplasts of photosynthetic tissues and in the plastids of roots and etiolated tissues of certain species. The relative proportions of these forms can vary within the organ of the same plant but also between species, notably in C_3 and C_4 plants (McNally et al. 1983; Cren and Hirel 1999). As such, it was proposed that each GS isoform might play a specific role, such as photorespiratory ammonium assimilation, nitrate reduction, N translocation and recycling depending on the organ, the developmental stage of the plant and the species examined (Cren and Hirel 1999; Miflin and Habash 2002; Martin et al. 2006;

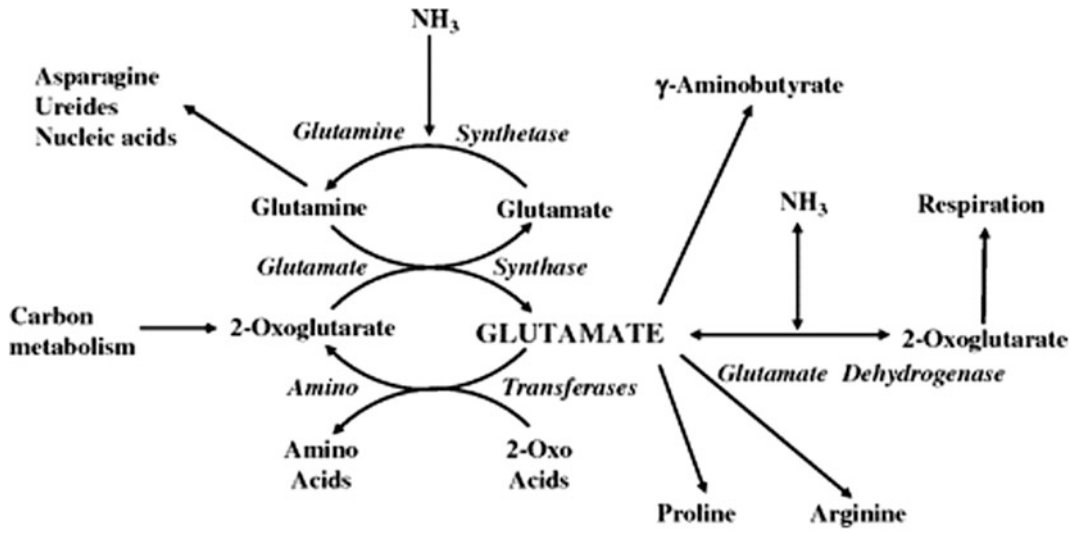

Fig. 16.1 Ammonium assimilation and synthesis of the main amino acids and N-containing molecules in plants

Bernard and Habash 2009). Moreover, it has been shown that in maize the up-regulation of some members of the GS multigene family may contribute to higher leaf N remobilization under drought conditions (Li et al. 2016) and in tilled soils when N availability is greater (Roach et al. 2016). Such findings indicate that the GS isoform complement can be finely tuned for optimal NUE depending on the environmental conditions.

C_4 plants such as maize exhibit a higher level of GS activity in the cytosol compared to the majority of C_3 plants, in which the enzyme activity is substantially lower (McNally et al. 1983). In maize, five different genes encoding GS1 and one encoding GS2 were originally identified (Li et al. 1993; Sakakibara et al. 1992) and further located on maize genetic maps (Gallais and Hirel 2004). The different GS isoenzymes are distributed between the bundle sheath and the mesophyll cells or expressed preferentially in the roots (Martin et al. 2006; Prinsi and Espen 2015). As in many other species, one of the five GS1 isoenzymes of maize is localized in the vascular tissue in which a high proportion of the protein is present in the phloem companion cells, where it is presumably involved in the transport of reduced N (Martin et al. 2006).

For maize, a putative role of GS in kernel yield has been proposed following a quantitative genetic approach, since quantitative trait loci (QTLs) for the leaf GS activity have been shown to be coincident with QTLs for kernel yield and its components. One QTL for thousand kernel weight (TKW) was coincident with the *gln3* locus (mapped by Hirel et al. 2001) and corresponding to the gene encoding cytosolic GS $Gln_{1\text{-}4}$, (Li et al. 1993) and two QTLs for thousand kernel weight and yield were coincident with the *gln4* locus (mapped by Hirel et al. 2001) and corresponding to the gene encoding cytosolic GS $Gln_{1\text{-}3}$, (Li et al. 1993). Further QTLs for GS gene loci have been identified in relation to remobilization of N from the leaf, stem and whole plant, post-anthesis N uptake (Gallais and Hirel 2004), and germination efficiency (Limami et al. 2002). More recently, a gene encoding the root GS1-1 isoenzyme was found to coincide with a QTL for low N tolerance (Luo et al. 2015a), strengthening the hypothesis that GS plays an important role in plant NUE.

Despite the wealth of information gathered over the last three decades on the regulation and function of GS with regard to plant N economy and the subsequent attempts to use the cytosolic

Table 16.1 Relationship between leaf glutamine synthetase activity and kernel yield in a panel of hybrids

[a]Yield components	[b]Low GS activity (1.20–1.46)	[b]High GS activity (2.34–2.58)	Increase (%)	[c]*p* value
Kernel number	460 ± 102	474 ± 104	3	0.11
Kernel yield	109 ± 35	120 ± 32	9.6	0.002
Thousand kernel weight	237 ± 53	255 ± 48	7.5	0.0007

Twenty-two different hybrids were pooled in two groups on the basis of total leaf GS activity. There were 11 hybrids in each group. The average value and standard error are given for yield and yield components for each of the two groups of hybrids. Three replicates harvested from three randomized blocks were used for each hybrid. The increase is the difference between the yield of hybrids exhibiting high GS activity and the yield of hybrids exhibiting low GS activity. The hybrids were produced by crossing 22 different lines (first name of the cross) with two commercial testers RAGT029 provided by RAGT Semences (Rodez, France) and EG001 provided by Euralis Semences (Lescar, France), (second name of the cross). The 11 hybrids exhibiting low GS activity were: EP1 × RAGT029, LO33 × RAGT029, ND30 × EG001, LO33 × EG001, CL18 × EG001, F473 × RAGT029, FC16 × EG001, F759 × LMGC0706, FV2 × RAGT029, FC209 × EG001, FC209 × RAGT029. The 11 hybrids exhibiting high GS activity are: H99 × EG001, FV232 × EG001, F7058 × EG001, CQ191 × EG001, Pa36 × RAGT029, NY302 × EG001, F712 × EG001, Commercial hybrid DK210 (RAGT Semences), F471 × EG001, Ia2132 × EG001, FV76 × LMGC0706. The experiment was performed in 2009 in an experimental field of the Institut National de la Recherche Agronomique at Versailles, France. Plants were grown on 170 kg N/ha, and the residual N provided by the soil was estimated to be approximately 60 kg/ha. Both phosphorus (P205) and potassium (K20) were also applied at 100 kg/ha. Plant density was 80,000/ha with a space of 80 cm between rows

[a]Kernel yield is expressed as g plant^{-1} and thousand kernel weight in g

[b]Total leaf glutamine synthetase (GS) activity is expressed as nmol min^{-1} mg^{-1} DW

[c]*p* value following a student t test

form of the enzyme as a target for plant improvement (Thomsen et al. 2014), only two studies performed on maize have demonstrated that GS1 overexpression improves kernel production. Transgenic plants that overexpress *Gln1-3* constitutively in the leaves exhibit an increase in kernel number, thus providing further evidence that the GS1.3 isoenzyme plays a major role in controlling kernel yield under high or low N fertilization conditions (Martin et al. 2006; Hirel et al. 2007b). In a more recent study, it was confirmed that in maize lines grown over two consecutive years in the field, overexpression of the two genes *Gln1-3* and *Gln1-4* improved yields and enhanced NUE (He et al. 2014). Several patents have been submitted, claiming that genes encoding GS expressed individually or simultaneously with tissue-specific promoters can be used for plant improvement (Abad et al 2015; Hirel and Perez 2016). However, it will be necessary to confirm that overexpression of one or several GS isoenzymes in hybrids, rather than lines per se, leads to an increase in kernel production. Performing field trials over multiple years at different locations (where maize is conventionally grown), as well as under stress conditions, (e.g., under drought or high population density), will also be required if commercial application is going to be successful. Encouragingly, preliminary studies have shown that grain yield is increased in hybrids exhibiting higher total leaf GS (Table 16.1).

Fd-GOGAT is a key enzyme mainly involved, in conjunction with plastidic GS (GS2), in the reassimilation of photorespiratory ammonium in leaves and stems. A pyridine nucleotide-dependent GOGAT isoenzyme (NADH-GOGAT; E.C. 1.4.1.14) is also present in higher plants. It is involved in the synthesis of glutamate, mostly in non-photosynthetic tissues, required to sustain plant growth and further development (Hirel and Lea 2001; Tabuchi et al. 2007; Plett et al. 2016).

Fd-GOGAT is encoded by a single gene (*ZmGOGAT1*, Plett et al. 2016) in maize and is mostly present in the bundle sheath cells (Becker et al. 1993) to recycle photorespiratory ammonium in the early stages of leaf development. In C_4 plants, the rate of photorespiration is much lower compared to C_3 plants (Bräutigam and Gowik 2016). A significant proportion of the

enzyme was also detected in the mesophyll cells allowing the recycling of glutamate in this cell type (Valadier et al. 2008). NADH-GOGAT is encoded by two distinct genes (*ZmGOGAT2 and 3*, Plett et al. 2016) in maize which seem to be involved in controlling other steps of inorganic nitrogen assimilation notably in the roots, since the enzyme activity increases in response to nitrate and ammonium N (Singh and Srivastava 1986) and at late stages of plant development (Plett et al. 2016). More recently, it has been reported that cadmium stress (Erdal and Turk 2016) and water lodging (Ren et al. 2017) induce a significant reduction in the activity of maize NADH-dependent GOGAT, thus leading to a reduction in plant NUE under abiotic stress conditions. Such findings confirm the important role of the pyridine nucleotide-dependent enzyme activity in maize N metabolism as previously demonstrated in another monocot, rice (Tabuchi et al. 2007). In a similar manner to GS, isoenzymes of GOGAT are key enzymes involved in inorganic N assimilation during the life cycle of maize, (Sakakibara et al. 1992; Plett et al. 2016), by virtue of their organ-specific, light-regulated, N- and stress-responsive modes of expression. However, currently, no attempts have been made to modulate the level of expression of GOGAT in a constitutive or organ-specific manner. Using reverse genetics, it will be interesting to establish whether as with rice, Fd-GOGAT (Zeng et al. 2017) and the two NADH-GOGAT isoenzymes (Yamaya and Kusano 2014) play a role in the N management of grain filling and if overexpression of the two types of GOGAT will improve kernel yield and kernel N content.

Another promising target for improving maize productivity is the reaction catalyzed by the enzyme glutamate dehydrogenase (NADH-GDH; E.C. 1.4.1.2), which has the potential capacity to assimilate inorganic N by combining ammonium with 2-oxoglutarate to form glutamate. Although GDH was originally thought to be the main source of glutamate synthesis in plants, an increasing number of studies have shown that GDH operates in the reverse direction of glutamate deamination to release organic acids when the cell is C-limited (Fontaine et al. 2012), even when the activity of the enzyme is increased in genetically modified plants (Labboun et al. 2009). Glutamate deamination could have both a metabolic and a signaling function (Tercé-Laforgue et al. 2004). Thus, GDH may be of importance under certain phases of plant growth and development when the translocation of C and N molecules is required, in line with the finding that the enzyme is confined to the companion cells of the phloem (Dubois et al. 2003; Tercé-Laforgue et al. 2004). It has been proposed that ammonium assimilation under certain physiological conditions can occur via NADH-GDH, which has an advantage over GS, as it uses less ATP per ammonia assimilated (Skopelitis et al. 2006, 2007). In maize, as in other plant species, two distinct genes encoding and GDH subunits have been identified so far (Hirel et al. 2005). In maize plants overexpressing the *E. coli* gene *gdhA* which encodes NADPH-GDH, the kernel biomass was higher than the controls when the plants were grown in the field under drought conditions (Lightfoot et al. 2007). This is in line with the finding that in maize, QTLs for GDH activity colocalized with QTLs for kernel yield (Dubois et al. 2003) and that in rice overexpressing a fungal GDH, grain yield was increased (Zhou et al. 2015). A patent filed in the USA by Schmidt and Miller (1999) described the use of several plant species, including maize, transformed with nucleotide sequences encoding the and subunits of *Chlorella sorokiniana* NADPH-GDH. These plants exhibited improved properties such as increased growth and enhanced osmotic stress tolerance. However, a more detailed physiological characterization of the maize plants overexpressing the *E. coli gdhA* gene and the *Chlorella* GDH subunits will be required to explain their improved performance. Moreover, as in tobacco, it would have been useful to establish whether the additional GDH activity was able to divert plant metabolism when the and subunits were overexpressed either individually or simultaneously (Tercé-Laforgue et al. 2013), thus rendering the transgenic plants more resistant to salt stress (Tercé-Laforgue et al. 2015).

16.3.3 Amino Acid Biosynthesis and Translocation

In comparison with the inorganic N uptake system and the ammonium assimilatory pathway, fewer studies have been devoted to the identification of the main steps of amino acid biosynthesis and interconversion in maize. These steps represent key elements in the control of grain filling, both in vegetative and reproductive organs (Seebauer et al. 2004; Cañas et al. 2009, 2011) and the production of grain protein (Lohaus et al. 1998; Uribelarrea et al. 2004). Nevertheless, it is clear that in a similar way to glutamine, the biosynthesis and translocation of other amino acids such as asparagine, aspartate, glutamate, and alanine have a major role in the control of assimilate supply and partitioning, both during vegetative growth and during grain development (Lea et al. 2007; McAllister et al. 2012; Trucillo Silva et al. 2017). Moreover, it has been recently shown that there is some genetic variability in the long-distance transport of amino acids in the phloem (Yesbergenova-Cuny et al. 2016). It is therefore likely that the concentration of amino acids and efficient translocation in the vascular tissue could be important factors in determining grain yield. These two factors are dependent upon N supply and N assimilation that both play an important role in the establishment and filling of the kernel through the modulation of the activity of enzymes involved in N and C metabolism (Singletary et al. 1990). There are several genes encoding enzymes involved in amino acid biosynthesis and interconversion differentially expressed in various organs or tissues of maize. This suggests that like those encoding GS and GOGAT they play specific roles in N assimilation and N recycling in not only vegetative (Todd et al. 2008; Plett et al. 2016) but also in reproductive organs (Cañas et al. 2009, 2010, 2011). More recently, a transcriptomic study highlighted the potential importance of two amino acid transporters that exhibited distinct expression patterns in the cob and the florets of maize suggesting that they have different roles in the mobilization of amino N in the ear (Pan et al. 2015).

Despite the importance of amino acid translocation from source leaves and amino acid interconversion in the developing ear during the grain-filling period, most of the studies aimed at understanding the genetic and physiological basis of NUE in maize have been largely focused on the vegetative source organs such as leaves and roots (Hirel et al. 2007b; Hirel and Gallais 2011). In contrast, fewer studies have been devoted to reproductive sink organs, although there is strong evidence that the developing ear (Seebauer et al. 2004; Liao et al. 2012) and the tassels (Pan et al. 2015) play major roles in the control of NUE, notably during the reproductive phase and over the kernel-filling period. In addition, it has been shown that N nutrition via the differential regulation of hormonal and C/N metabolism in the developing ear is a key determinant of kernel production (Liao et al. 2012). During this period, the translocation of glutamate, glutamine, and aspartate to the ear and the subsequent synthesis of asparagine during the transfer of organic N from the cob to developing kernels are important metabolic processes that probably condition kernel-filling efficiency (Seebauer et al. 2004; Cañas et al. 2010). It can be assumed that primary N assimilation and N remobilization occur simultaneously in source leaves and contribute almost equally to the supply of organic N to the developing ear (Gallais et al. 2007; Hirel et al. 2007b). Thus, it is likely that in maize the supply of N assimilates, their efficient translocation, and their conversion during kernel formation all contribute to the overall plant NUE. Unfortunately, our knowledge of the coordinated regulation of these three metabolic processes remains limited, notably if we consider that there is a competition between the tassel and the developing ear for N partitioning which can be detrimental for grain production when N is limiting (Fox et al. 2017). In parallel, further research will be required to study the molecular physiology of plant N translocation during kernel

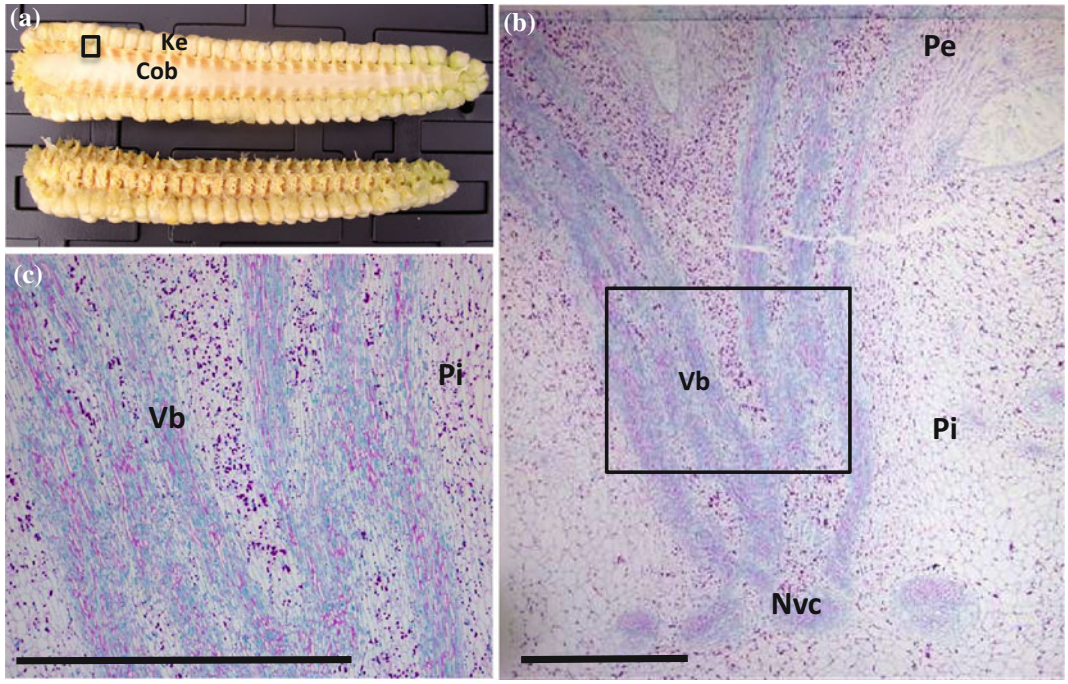

Fig. 16.2 Complex structure of the maize ear vasculature. **a** Structure of the ear 14 days after silking showing the connection between the kernels (Ke) and the cob (Cob). The black square indicates the position of the fragment of tissue used for microscopy. **b** Structural organization of the connection between the kernel and the cob. **c** Magnification of the black square shown in (**b**). Nvc: nodal vascular complex; Pe: pedicel; Pi: pith region of the cob; Vb: main vascular bundle. Thin tissue sections (1 µm) were stained by the periodic acid–naphthol blue black (PAS–NBB) method as described by Sangwan et al. (1992). Bars = 100 µm

formation and to identify the key structural and regulatory elements involved, taking into account the complex cellular and vascular structure of the connection between the cob and the kernels (Fig. 16.2). For example, the finding that in the pedicel a specific GS isoenzyme is actively synthesized at later stages of ear development indicates that glutamine synthesis and thus N remobilization are important within the tissues connecting the cob and the developing kernels, during grain fill (Muhitch 2003). Moreover, determining whether the genetic variability for translocation and conversion of N metabolites has an impact not only on storage protein deposition but also on the export and storage of C during the kernel-filling process is also worthy of investigation (Seebauer et al. 2010).

For this reason, studies were undertaken to characterize at both biochemical and molecular levels, the main steps of N metabolism during the development of the maize ear. These investigations were carried out either in maize lines exhibiting contrasting NUE or in GS-deficient mutants, in which glutamine production in source leaves was reduced (Cañas et al. 2009, 2010, 2011). On the basis of gene expression, enzyme activity, and metabolic profiling studies, it has been shown that both in the cob and in the developing kernels, the maize lines and mutants have acquired different strategies for N management depending on the level of N fertilization (Cañas et al. 2009, 2010, 2011). Such findings have prompted researchers to further investigate which components could be important in the control of grain filling during ear development, taking advantage of the genetic diversity for NUE in maize. Quantitative genetic studies were thus undertaken to identify QTLs for the main

steps of N assimilation in the developing ear of maize. These included investigations as to whether there were colocalizations with QTLs for cob morphological traits, kernel yield, and putative candidate genes in order to identify metabolic pathways and regulatory functions putatively involved in the determination of yield (Cañas et al. 2012; Jansen et al. 2015).

During the grain-filling period, the changes in physiological traits were monitored in the cob and in the developing kernels, representative of C and N metabolism in the developing ear. The correlations between these physiological traits and traits related to yield were examined and localized with the corresponding QTLs on a genetic map. Unexpectedly, glycine and serine metabolism in developing kernels and the cognate genes appeared to be of major importance for kernel production (Cañas et al. 2012). Nevertheless, the importance of kernel glutamine synthesis in the determination of yield (Limami et al. 2002) and kernel amino acid content (Deng et al. 2017) was also confirmed (Cañas et al. 2012). More recently, single-nucleotide polymorphisms (SNPs) and candidate genes encoding asparaginase, asparagine synthetase, and glutamine synthetase were found to be associated with kernel amino acid traits, providing further insights into the genetic basis of amino acid biosynthesis and degradation in maize kernels (Deng et al. 2017). Such quantitative and association genetic approaches provided the groundwork for improving our understanding of the genetic and physiological bases of N metabolism in the developing ear (see paragraph 5). However, as for the recent study performed on leaves (Cañas et al. 2017), further work will be required to extend the study to maize germplasm exhibiting a wider genetic diversity and using more integrated systems biology approaches. Finding mutants and producing transgenic plants in which the modifications of assimilate partitioning and translocation are specifically targeted will also be required. This implies that large-scale "omics"-based phenotyping protocols, together with the appropriate bioinformatics/modeling tools, are available to fully exploit the large data sets generated from such studies (Simons et al. 2014b; Sweetlove et al. 2017).

16.3.4 Regulation of Nitrate Signaling and Assimilation

Maize Dof1 is a member of the DNA binding with one finger (Dof) family of transcription factors unique to plants and is an activator of the expression of a range of genes associated with organic acid metabolism (Yanagisawa 2004). Transgenic Arabidopsis expressing Dof1 under the control of a maize pyruvate phosphate dikinase (PPDK) promoter exhibited a remarkable elevation in the concentration of amino acids, especially glutamine, and increased growth under low nitrogen conditions (Yanagisawa et al. 2004). Although in an investigation of Cavalar et al. (2007), it was concluded that the Dof1 transcription factor does not play a major role in the control of N or C metabolism in maize. However, a recent study has shown that overexpressing the maize Dof1 transcription factor in wheat leads to an improvement in growth and productivity (Peña et al. 2017). This indicates that the Dof1 transcription factor may be an interesting candidate to boost agronomic performances of cereals.

Other regulatory genes such as *NLP7* are central components in regulating nitrate response in Arabidopsis (Castaings et al. 2009). Nine *NLP* genes were recently identified in maize. Among them, *ZmNLP6* and *ZmNLP8* regulate nitrate signaling in Arabidopsis and were able to increase plant biomass and yield when overexpressed in the model species (Cao et al. 2017). Such a finding strengthens the idea that more work is required to assess the role of regulatory genes in crop productivity in general and in maize in particular, both using genetic manipulation and association genetics.

16.4 Toward More Integrated Genetic and Physiological Approaches for NUE Improvement

16.4.1 "Omics"-Based Approaches and Their Integration

Due to the complexity of both the biological systems involved in the control of NUE and their regulation at the cellular, organ, or whole-plant levels, integrated studies, including systems biology and genome-scale metabolic approaches have been developed. These include the available transcriptomic and proteomic profiles along with metabolomic data sets and enzyme activity profiles (Fukushima and Kusano 2014; Simons et al. 2014b).

Several transcriptome studies were originally carried out to evaluate modifications in gene expression under low and high N conditions developed in order to identify genome-wide transcriptional circuits in various organs and tissues during maize development (Amiour et al. 2012; Plett et al. 2015) and particularly those related to N-responsive genes. A number of these studies have highlighted the complexity of the regulatory mechanisms involved in the transcriptional control of leaf or root gene expression under N-limiting and non-limiting conditions (Amiour et al. 2012; Luo et al. 2015a; Trevisan et al. 2015) and in combination with other abiotic stresses such as drought (Humbert et al. 2013). Depending both on the duration and intensity of the N-limiting stress applied, most of the studies in maize have ended up with a portfolio of genes involved in a variety of developmental, metabolic, and regulatory functions (Amiour et al. 2012; Humbert et al. 2013) that also include transcription factors (Chen et al. 2015). The biological processes that were the most significantly reduced in leaves under N-limiting conditions were those involved in photosynthetic C and N metabolism. These were followed by various other metabolic processes involved, for example, in metal homeostasis/detoxification, plant pigment synthesis, phospholipid synthesis, and response to stress (Amiour et al. 2012; Humbert et al. 2013). In roots, the reductions in the level of transcript for N transport, N reduction, and N assimilation were generally predominant and different to that found for the leaves, indicating that there is an organ-specific, transcriptional regulation in response to N deficiency (Zanin et al. 2015; Plett et al. 2015).

In maize, transcriptome studies of GS mutants deficient in cytosolic GS (GS1) have also revealed deficiencies in key reactions or key regulatory proteins involved in primary N metabolism (Amiour et al. 2014). Interestingly, distinct transcriptome response signatures for urea- and nitrate-supplied maize plants have been identified. The data indicated that there is a urea- and a nitrate-specific pattern of root gene expression and that the concomitant presence of urea and nitrate enhanced the expression of genes encoding proteins and enzymes involved in nitrate uptake and reduction (Zanin et al. 2015). In another study, Guo et al. (2014) showed that the expression of genes involved in N remobilization was enhanced when maize plants were grown under organic N supply. Such observations suggest that the regulation of N-responsive gene expression under agronomic conditions depends on the type of fertilization used and is probably more complex than that occurring in plants grown under controlled conditions on a single N source. In addition, it will be necessary to consider that the transcriptional regulatory mechanisms triggered during N limitation will vary not only in different inbred lines exhibiting contrasted NUE (Zamboni et al. 2014) but also in their parental lines as well as wild ancestors and ancient genotypes which are often adapted to N deficiency (Bi et al. 2014; Han et al. 2015a). It will thus be essential to take into account these differences if genes are to be used as putative markers to improve NUE and maize production. Nevertheless, a number of common N-responsive genes were found in different genotypes including hybrids under both controlled and field growth conditions. This finding led Yang et al. (2011) to propose that a small set of N-responsive genes could be used as biomarkers to monitor the N status of maize, notably in the leaves. A number of these genes were also found

in the study of Amiour et al. (2012), thus strengthening the idea that they could be used for quantitative genetic studies for detecting NUE QTLs (Chen et al. 2015). Developing arrays containing these NUE marker genes, like those designed for drought and kernel oil content (Xu et al. 2017), could provide an interesting perspective for marker-assisted breeding. However, the exact function of a number of these genes still needs to be identified as they probably encode specialized metabolic functions that have proliferated to a much greater degree during plant evolution and crop domestication, compared to their primary metabolism counterparts (Chae et al. 2014).

In parallel with classical genome-wide transcriptional profiling, the occurrence of microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) has been investigated in plants grown under different levels of N supply. This has led a number of research groups to investigate if epigenetic regulation (Sirohi et al. 2016) could be an important component of the response of maize to a N stress (Zhao et al. 2012; Lv et al. 2016). Such a putative regulatory function mediated by the action of noncoding RNAs was highlighted by the finding that significant differences in their accumulation were observed according to the level of N nutrition, as well as their spatiotemporal expression pattern in maize root and leaf tissues (Trevisan et al. 2012; Zhao et al. 2012). This provided new insights into the timing and tissue specificity of transcriptional regulation under low N nutritional conditions. The genes targeted by miRNAs have various and ubiquitous functions, encompassing a variety of developmental and metabolic processes that are not necessarily directly linked to NUE (Xu et al. 2011). It has been proposed that the genetic manipulation of the expression of miRNAs could be an alternative method of improving NUE in crops (Fischer et al. 2013; Sinha et al. 2015). In line with this proposal, manipulation of microRNA528 (miR528), which is a conserved monocot-specific small RNA, appears to be promising for improving tolerance to N deficiency (Yuan et al. 2015). Other forms of inherited epigenetic memory induced by the plant N status were found in rice leaves (Kou et al. 2011) and Arabidopsis roots (Widiez et al. 2011), as there were locus-specific altered levels of DNA methylation. It will thus be interesting to investigate if such additional epigenetic control mechanisms occur in maize, since in rice 50% of the altered methylation patterns were recaptured in the progeny which eventually led to a higher tolerance to N starvation (Kou et al. 2011).

As a complement to genome-wide transcriptome studies, proteome studies are able to provide additional information on the quantity of expressed proteins and their posttranslational modifications such as phosphorylation and glycosylation under various abiotic stress conditions including N deficiency (Hu et al. 2015; Nouri et al. 2016). In maize, an increasing number of studies have been carried out in order to examine changes in the root and shoot protein profile using plants grown either under control conditions or in the field under high and low N supply (Amiour et al. 2012; Jin et al. 2015; Trevisan et al. 2015; Wei et al. 2016). Results from these studies showed that the amounts of enzyme proteins that have a pivotal role in C and N metabolism were higher when plants were fed with nitrate. Many other proteins involved in maintaining the energy and redox status of the cell and signal transduction were also shown to be N-responsive (Table 16.2). Other proteomic studies performed on the developing ear showed that under N-limiting conditions, during which kernel yield was markedly reduced, stress-responsive proteins and proteins involved in hormone metabolism and function were preferentially accumulated (Liao et al. 2012). Such observations indicate that in reproductive and vegetative organs, different functional classes of proteins are involved in response to N deficiency. Differential expression of proteins was also observed in low-N-sensitive and low-N-tolerant maize genotypes in response to various N treatments (Nazir et al. 2016). Such findings, on top of improving our knowledge of the response of maize to low N supply, suggest that protein profiling could be a way to select for genotypes that perform well when N fertilization is reduced. This information can be used for quantitative

Table 16.2 List of proteins showing changes in their level of accumulation under low and high nitrogen nutrition regimes in seven different studies

[a]Protein name	Functional category	Sequence number	[b]References
Phosphoribulokinase	CO_2 fixation	GenBank: ACG34613.1	1, 2, 6
Ribulose-1,5-bisphosphate carboxylase	Carbon metabolism	NCBI: NP_043033.1	1, 2
NADP–malate dehydrogenase	Carbon metabolism	NCBI: NP_001105603.1	2, 4, 5, 6
α−1,4 glucan-protein synthase	Cell wall synthesis	NCBI: NP_001105598.1	1, 2
Glycine-rich protein	Cell wall synthesis	GenBank: ACG38893.1	1, 2, 3
Ferredoxin	Energy transfer	NCBI: NP_001336742.1	1, 6
ATP synthase	Energy transfer	NCBI: NP_001148315.1	1, 2, 3, 4, 5
Enolase	Glycolysis	GenBank: CAA39454	1, 2, 4
Glutamine synthetase	Nitrogen metabolism	NCBI: NP_001105538.1	1, 2, 4, 6,7
Oxygen-evolving enhancer protein	Photosynthesis	GenBank: ACG31595.1	1, 2, 4
Chlorophyll *a-b* binding protein	Photosynthesis	GenBank: ACG28457.1	1, 2, 6
Arginine decarboxylase	Polyamine biosynthesis	GenBank: ONM11577.1	1, 6
Cysteine protease	Proteolysis	GenBank: ACG36262.1	1, 2, 3
Eukaryotic translation initiation factor 2A	Signaling	Swiss-Prot: P80639.2	1, 2
Peroxiredoxin	Stress response	GenBank: ACG24946.1	1, 3

[a]Differentially expressed proteins identified in different organs of maize under low nitrogen supply
[b]Reference used for the comparison: Amiour et al. (2012) (1), Jin et al. (2015) (2), Liao et al. (2012) (3), Nazir et al. (2016) (4), Trevisan et al. (2015) (5), Wei et al. (2016) (6), Yan et al. (2014) (7)

genetic studies and proteome-based marker-assisted selection (Hu et al. 2015). Identification of these genotypes based on proteomic studies should be preferably performed in the field, since under these growth conditions the impact of cropping systems and thus of the rhizosphere in promoting NUE can also be assessed (Yan et al. 2014).

As for the other "omics"-based studies, an increasing number of metabolomic investigations have been carried out using both model and crop plants, with the aim of identifying changes in metabolite concentrations under various biotic and abiotic stresses including N deficiency (Fukushima and Kusano 2014; Simons et al. 2014b). These metabolic profiling studies have also been valuable in improving our understanding of the interactions between C and N metabolism and their interaction with other stresses such as drought and phosphorus deficiency. Such approaches have allowed the identification of new compounds that accumulate in response to a given stress, as well as those sharing a common pattern of accumulation across various stress conditions (Schlüter et al. 2013; Sun et al. 2016). In addition, a number of plant metabolic databases are now available that will facilitate the development of plant systems biology using combined "omics"-based approaches (Colmsee et al. 2012; Fukushima and Kusano 2013).

When examining the plant metabolic profile following N deprivation, a general decrease in most of the metabolites involved in both C and N primary assimilation is observed which could partly explain why either vegetative biomass or kernel production is reduced. Interestingly, in a number of studies performed on maize it was also observed that the accumulation of secondary metabolites particularly those used as precursors of cell wall synthesis is an important indicator of N deficiency (Amiour et al. 2012; Schlüter et al. 2013) and the reduction of plant productivity (De Abreu e Lima et al. 2017; Cañas et al. 2017). This is in line with the investigations into maize GS-deficient mutants in which grain yield was

strongly reduced (Broyard et al. 2009; Amiour et al. 2014). Thus, N deficiency, or a perturbation of primary N assimilation, has a strong impact on maize growth and development through the altered synthesis of metabolites used as the precursors required for lignin and cellulose production. Changes in the content of a number of unknown metabolites or metabolites of unknown function, notably chlorogenates, were frequently observed when maize plants were grown under N-limiting conditions. Such findings indicate that more in-depth chemical and biological investigations need to be carried out before they are identified (Nakabayashi and Saito 2015).

Such a metabolic snapshot representative of the impact of N deficiency can then be used alone or together with the other "omics" data sets to identify targets (genes, proteins, enzyme activities, and metabolic pathways) for improving biomass or kernel production (Lisec et al. 2011; Amiour et al. 2012, 2014; Riedelsheimer et al. 2012a; Beatty et al. 2016).

Although not related to NUE, it has recently been shown that metabolic profiles at an early stage of plant development can constitute a reliable tool for predicting maize hybrid performance in the field, strengthening the potential of the technique for breeding purposes (De Abreu e Lima et al. 2017). In another recent study, leaf metabolic profiling in combination with enzyme activity profiling was used to link leaf physiology to kernel yield in genetically distant maize lines (Cañas et al. 2017). In addition, genome-wide metabolic profiling studies have in certain cases helped to link quantitative traits to their causal genetic loci and key metabolic regulators. Whole-genome and metabolic prediction models were built by fitting in the effects of all single-nucleotide polymorphisms (SNPs) present in the genome and the accumulation of metabolites. This allowed a reliable screening of large collections of diverse inbred lines for their potential to create superior hybrids (Riedelsheimer et al. 2012a). Altogether, these studies strongly suggest that metabolomic tools are likely to be the most appropriate for the prediction of plant performance.

16.4.2 "Omics" Integration, Metabolic Modeling, and Integrative Analysis of Regulatory Networks

In the majority of the "omics"-based studies in which the response of a plant to N limitation was investigated, there was no direct relationship between transcriptome, proteome, or metabolome information. However, the data suggested that at least in maize, the main plant metabolic functions that were altered as a result of N deficiency were conserved across the different "omics" (Amiour et al. 2012, 2014). A limited correspondence of 20% between the protein profiles and the accumulation of the corresponding transcripts was also observed in maize roots (Trevisan et al. 2015). In other studies, only a few correlations between changes in amino acids and transcripts encoding enzymes involved in amino acid biosynthesis were identified (Schlüter et al. 2013). This is in agreement with the finding that mRNA levels were generally more N-responsive than the corresponding enzyme activities (Plett et al. 2016). Such findings suggest that transcript abundance is not necessarily the main factor that regulates the final corresponding enzyme activity. This is likely because complex and still uncharacterized network interactions are probably occurring between gene transcription, regulatory proteins, catalytically active proteins, and metabolite accumulation (Fernie and Stitt 2012). This would suggest that other regulatory elements such as uncharacterized genes or metabolites could have important functions within the biological networks involved (Urano et al. 2010). Moreover, it is generally admitted that "omics" studies only provide a narrow and static picture of the physiological status of a given organ, at a particular stage of plant development (Fernie and Stitt 2012). Thus, additional fluxomic studies based on the use of ^{15}N- and ^{13}C-labeled compounds and network flux analyses, similar to those performed by Cañas et al. (2017), are required. Such studies should be able to determine if there are links between metabolite accumulation, and the

metabolic fluxes occurring in mutants, genetically modified crops or genotypes exhibiting contrasting NUE under low and high N supply (Kruger and Ratcliffe 2012, 2015; Shachar-Hill 2013; Freund and Hegeman 2017). The flow of labeled metabolites determined experimentally can also be predicted by flux balance analysis (FBA) in metabolic models (Junker 2014; Srivastava et al. 2016). Such a predictive value of FBA is illustrated in a recent study in which a maize leaf genome-scale metabolic (GSM) model was used to corroborate the hypotheses based on the results of a ^{15}N-labeling experiment and enzyme activity measurements determined in vitro (Cañas et al. 2017). In agreement with the ^{15}N-labeling experiments, the in silico results indicated that the metabolic fluxes going through the photorespiratory glycolate pathway and the ammonia assimilatory pathway could be very different in genetically distant maize lines.

In addition, GSM models could be an interesting approach not only for the integration of "omics" data but also to identify putative candidate genes, proteins, and metabolic pathways contributing to plant growth and development as a function of N availability (Simons et al. 2014b; Kruger and Ratcliffe 2015). A GSM model for the maize leaf was created integrating C_4 carbon fixation, thus allowing an investigation into N assimilation by modeling the metabolic fluxes and interactions within the leaf (Simons et al. 2014a). Available "omics" data (Amiour et al. 2012, 2014) were also used to introduce regulatory constraints in the model to simulate a N limited condition, along with the *gln$_{1\text{-}3}$ and gln$_{1\text{-}4}$* mutants deficient in isoenzymes of glutamine synthetase (Martin et al. 2006). Such leaf models, with the addition of other maize organ-specific models, can be integrated into a whole-plant GSM for maize. Such a model of maize will help to elucidate the flow of N from the root to the other tissues in the plant, from the shoot to the ear, and within the developing ear (Cañas et al. 2010). By modeling the entire plant, non-intuitive bottlenecks in N metabolism can be determined, which then can be used to suggest genetic interventions through mutagenesis, transgenic technology, or marker-assisted selection to increase NUE and plant productivity. For example, the flow of sugars and N to the kernel tissue can be analyzed to guide the increase of carbohydrate and protein content. Since maize is also used for cellulosic biofuels, the whole-plant GSM could be used as a predictive tool to propose genetic disruptions that reduce the lignin content without affecting the mechanical integrity of the plant (Simons et al. 2014a).

Following the various "omics"-based studies, the data sets necessary to develop systems biology have been increasingly used to establish if relationships exist between mRNA, metabolite and protein accumulation, enzyme activities, and phenotypic traits related to plant growth and development both under optimal growth (Farré et al. 2015; Wuyts et al. 2015) and abiotic stress conditions (Obata et al. 2015). However, in these studies, detailed interpretation of the underlying physiology, either in terms of metabolite accumulation or metabolic fluxes, was not fully exploited. This was often due to the lack of experimental data in the field and in some cases to appropriate bioinformatics tools (Shen et al. 2013). Nevertheless, these studies have opened up interesting perspectives concerning the use of "omics"-assisted breeding techniques for narrowing the genotype/phenotype gap of complex traits, such as yield and biomass production both when primary metabolism (Gehan et al. 2015; Tohge et al. 2015) and secondary metabolism (Schilmiller et al. 2012) are considered.

More recently to fill the gap between the genotype and phenotype of complex traits, new approaches combining "omics" data have taken advantage of the worldwide genetic diversity of maize lines and inbred populations (Wen et al. 2015; Toubiana et al. 2016; Cañas et al. 2017). The aim of such approaches has been to determine if in a metabolic network in which metabolites and enzyme activities are interconnected, whether components of the metabolic network could be used as selection markers for breeding maize with a superior agronomic performance. All three studies revealed the important role of central metabolism including carbohydrate and amino acid metabolism and the

tricarboxylic acid cycle in both the connectivity of the network and in plant performance.

In the study by Cañas et al. (2017) in which metabolomic, biochemical, fluxomic, and metabolic modeling approaches were combined, both correlation studies with yield-related traits and metabolic network analyses allowed the description of a maize ideotype with a high grain yield potential. Such an ideotype is characterized by a low accumulation of soluble amino acids and carbohydrates in the leaves and a high activity of enzymes involved in the C_4 photosynthetic pathway and in the synthesis of amino acids derived from glutamate. It was thus proposed that a number of metabolites and enzyme activities could be used as physiological markers for breeding purposes, including genome-wide association genetics and marker-assisted selection. Such a study perfectly illustrates the need to improve our understanding of the intricate network of physiological and phenotypic traits and how they respond when there are environmental changes such as N deficiency (Poorter et al. 2013).

16.5 Genome-Wide Association Genetics and Marker-Assisted Selection

In most of the studies in which QTL mapping was performed using recombinant inbred lines (RILs) to identify the genetic basis of NUE, the size of the characterized chromosomal regions was generally between 5 and 30 cm depending on the size of the population of RILs and of the measured agronomic and phenotypic trait (Gallais and Hirel 2004; Jansen et al. 2015). In addition, fine mapping and ultimately candidate gene identification can be carried out by positional cloning (Gallavoti and Whipple 2015). However, in rather large chromosomal regions, this is generally limited by the production of a sufficient number of informative recombinations and by a low genetic diversity of the traits of interest (Salvi and Tuberosa 2007). Currently, narrowing down regions in the 100 kb range (Raihan et al. 2016) and map-based cloning of the gene of interest (Gallavoti and Whipple 2015) is becoming more and more accessible (Singh et al. 2017). Recently, map-based cloning was successfully used to identify homeologous genes involved in NUE in tobacco (Edwards et al. 2017). Although such an approach was not successfully used to identify genes underlying NUE traits in maize, the recent production of balanced multiparental (MAGIC) populations (Dell'Acqua et al. 2015) and of ultra-high-density maps (Liu et al. 2015b; Su et al. 2017) should provide powerful tools leading to higher power and definition in QTL mapping for complex traits such as NUE.

Association mapping studies, which are based on linkage disequilibrium (LD), also allow the identification of candidate genes underlying the identified QTLs. However, the success of gene-based association studies largely depends on the candidate gene(s) chosen for a specific phenotypic trait, particularly in maize (Gallavoti and Whipple 2015). The choice of the genes can be made taking into account the results previously obtained using a classical QTL approach, followed by functional validation of some members of the multigene families encoding enzymes involved in ammonia assimilation and amino acid biosynthesis, such as GS (Martin et al. 2006).

Recently, leaf metabolite profiling techniques have been used successfully to dissect complex traits in maize through the use of genome-wide association mapping both in maize lines (Riedelsheimer et al. 2012a; Zhang et al. 2015) and hybrids (Riedelsheimer et al. 2012b). The study of Zhang et al. (2015) showed that the fine-tuning of the expression of genes that drive the C_4 carbon shuttle are likely key determinants of yield. The levels of central metabolites in C and N metabolism are determined by genetic variation in key genes involved in CO_2 capture such as carbonic anhydrase and further movement of C-containing molecules such as malate.

Thus, metabolome-assisted breeding techniques, in addition to genome-assisted selection of superior hybrids, would appear to be promising methods of narrowing the genotype/phenotype

gap of complex traits such as NUE (Luo 2015b). In addition, they would be useful for identifying domestication-selected genes controlling NUE and other important agronomic traits (Wallace et al. 2014). Narrowing the genotype/phenotype gap through association analysis will also be possible for root traits of plants grown under varying concentrations of N. This will allow both the identification of genes and the development of functional markers for the selection of maize lines with improved root architecture and yield under N-deficiency stress conditions (Abdel-Ghani et al. 2015).

16.6 Conclusions and Future Challenges

Together with the existing knowledge of the whole-plant physiology of maize, several sets of "omics" data related to the plant response to short-term and long-term N deficiency are now available. With the development of more powerful computing and genetic approaches including the construction of a GSM, significant progress has been made in integrating the "omics" data and linking them to a plant phenotype in terms of NUE, plant biomass production, and yield.

Although there is often a limited correlation between differences in the accumulation of metabolites, proteins, and mRNA due to the complexity inherent in biological networks, computational and modeling approaches integrating "omics" data have identified the key biological components involved in the NUE of maize.

These integrated studies showed that, depending on the level of N nutrition, major biological functions such as C assimilation, several metabolic pathways, and stress responsive and regulatory elements linked to C utilization share common characteristics across the different "omics" in a given organ or tissue.

Metabolomic and ^{15}N-labeling fluxomic studies confirmed that specific steps across the "omics" are limiting during vegetative growth and during the grain-filling period.

The recent achievements in quantitative genetic and association genetic studies, integrating "omics" data, have identified a number of key loci involved in plant productivity with respect to N uptake and utilization, thus linking a phenotype to DNA markers and ultimately to a gene or a set of genes. However, the functional validation of these structural or regulatory genes using transgenic technologies, mutagenesis, or by studying the relationship between allelic polymorphisms and the trait of interest either at a single gene or genome-wide level was up until to now, rarely successful. To observe a positive effect of overexpression or the impairment of a particular gene or a group of genes, the transformation or the selection of the mutant needs to be conducted in a genetic background most suited for observing the desired phenotypic effect. Moreover, if the individual level of expression of stacks of genes has to be tested in a particular organ or tissue, it will be necessary to observe an impact on plant NUE and ultimately on its productivity. Whole-genome association genetic studies need to be developed further to link NUE, NUE-related traits, and "omics" markers related to NUE, to DNA markers for future marker-assisted breeding strategies.

The new gene, protein, and metabolite markers identified thanks to the knowledge gained from the "omics studies" performed on different genotypes under various environmental conditions could be used for the production of diagnostic tools. Such tools will allow the monitoring of their levels of expression, regardless of the genetic background or of undesirable environmental effects. These diagnostic tools, if easy to use and affordable, could be utilized by breeders to screen for the best-performing genotypes under low and high N fertilizer inputs.

References

Abad MS, Coffin M, Goldman BS (2015) Genes encoding glutamine synthetase and used for plant improvement. US patent 20150184189 A1

Abdel-Ghani AH, Kumar B, Pace J, Jansen C, Gonzalez-Portillla PJ, Reyes-Matamoros J, San Martin JP, Lee M, Lübberstedt T (2015) Association

analysis of genes involved in maize (*Zea mays* L.) root development with seedlings and agronomic traits under contrasting nitrogen levels. Plant Mol Biol 88:133–147

Allen SM, Guo M, Loussaert DF, Rupe M, Wang H (2014) Enhanced nitrate uptake and nitrate translocation by over-expressing maize functional low-affinity nitrate transporters in transgenic maize. US patent 20160010101 A1

Amiour N, Imbaud S, Clement G, Agier N, Zivy M, Valot B, Balliau T, Armengaud P, Quilleré I, Cañas RA, Tercé-Laforgue T, Hirel B (2012) The use of metabolomics integrated with transcriptomic and proteomic studies for identifying key steps involved in the control of nitrogen metabolism in crops such as maize. J Exp Bot 63:5017–5033

Amiour N, Imbaud S, Clément G, Agier N, Zivy M, Valot B, Balliau T, Quilleré I, Tercé-Laforgue T, Dargel-Graffin C, Hirel B (2014) An integrated "omics" approach to the characterization of maize (*Zea mays* L.) mutants deficient in the expression of two genes encoding cytosolic glutamine synthetase. BMC Genom 15:1005

Andrews M, Lea PJ, Raven JA, Lindsey K (2004) Can genetic manipulation of plant nitrogen assimilation enzymes result in increased crop yield and greater N-use efficiency? An assessment. Ann Appl Biol 145:25–40

Andrews M, Raven JA, Lea PJ (2013) Do plants need nitrate? The mechanisms by which nitrogen form affects plants. Ann Appl Biol 163:174–199

Bashan Y, de-Bashan LE, Prabhu, SR, Hernandez JP (2014) Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). Plant Soil 378:1–33

Beatty PH, Klein MS, Fisher JJ, Lewis IA, Muench DG, Good AG (2016) Understanding plant nitrogen metabolism through metabolomics and computational approaches. Plants 5:39. https://doi.org/10.3390/plants5040039

Becker TW, Perrot-Rechenman C, Suzuki A, Hirel B (1993) Subcellular and immunocytochemical localization of the enzymes involved in ammonia assimilation in mesophyll and bundle sheath strands of maize leaves. Planta 191:129–136

Bernard SM, Habash DZ (2009) The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. New Phytol 182:608–620

Bi YM, Meyer A, Downs GS, Shi X, El-kereamy A, Lukens L, Rothsten SJ (2014) High throughput RNA sequencing of a hybrid maize and its parents shows different mechanisms responsive to nitrogen limitation. BMC Genomics 15:77

Binder DL, Sander DH, Walters DT (2000) Maize response to time of nitrogen application as affected by level of nitrogen deficiency. Agron J 92:1228–1236

Bloom AJ (2015) The increasing importance of distinguishing among plant nitrogen sources. Curr Opin Plant Biol 25:10–16

Bräutigam A, Gowik U (2016) Photorespiration connects C_3 and C_4 photosynthesis. J Exp Bot 67:2953–2962

Broyard C, Fontaine JX, Molinie R, Cailleu D, Tercé-Laforgue T, Dubois F, Hirel B, Dubois F, Mesnard F (2009) Metabolomic profiling of two cytosolic glutamine synthetase maize mutants (*Zea mays* L.) using 1H-nuclear magnetic resonance. Phytochem Anal 21:102–109

Cameron KC, Di HJ, Moir JL (2013) Nitrogen losses from the soil/plant system: a review. Ann Appl Biol 162:145–173

Cañas RA, Quilleré I, Christ A, Hirel B (2009) Nitrogen metabolism in the developing ear of maize (*Zea mays* L.): analysis of two lines contrasting in their mode of nitrogen management. New Phytol 184:340–352

Cañas RA, Quilleré I, Lea PJ, Hirel B (2010) Analysis of amino acid metabolism in the ear of maize mutants deficient in two cytosolic glutamine synthetase isoenzymes highlights the importance of asparagine for nitrogen translocation within sink organs. Plant Biotech J 8:966–978

Cañas RA, Amiour N, Quilleré I, Hirel B (2011) An integrated statistical analysis of the genetic variability of nitrogen metabolism in the ear of three maize inbred lines (*Zea mays* L). J Exp Bot 62:2309–2318

Cañas RA, Quilleré I, Gallais A, Hirel B (2012) Can genetic variability for nitrogen metabolism in the developing ear of maize be exploited to improve yield? New Phytol 194:440–452

Cañas RA, Yesbergenova-Cuny Z, Simons M, Chardon F, Armengaud P, Quilleré I, Cukier C, Gibon G, Limami AM, Nicolas S, Brulé L, Lea PJ, Maranas CD, Hirel B (2017) Exploiting the genetic diversity of maize using a combined metabolomic, enzyme activity profiling, and metabolic modeling approach to link leaf physiology to kernel yield. Plant Cell 29:919–943

Cao H, Qi S, Sun M, Li Z, Yang Y, Crawford NM, Yong W (2017) Overexpression of the maize *ZmNLP6* and *ZmNLP8* can complement the *Arabidopsis* nitrate regulatory mutant nlp7 by restoring nitrate signaling and assimilation. Front Plant Sci 8:1703

Cassán F, Diaz-Zorita M (2016) *Azospirillum* sp. in current agriculture: from the laboratory to the field. Soil Biol Biochem 103:117–130

Castaing L, Camargo A, Pocholle D, Gaudon V, Texier Y, Boutet-Mercier S, Taconat L, Renou JP, Daniel-Vedele F, Fernandez E, Meyer C, Krapp A (2009) The nodule interception-like protein 7 modulates nitrate sensing and metabolism in maize. Plant J 57:426–435

Cavalar M, Phlippen Y, Kreuzaler F, Peterhaensel C (2007) A drastic reduction in DOF1 transcript levels

does not affect C4 specific gene expression in maize. J Plant Physiol 164:1665–1674
Ceccarelli S (2014) GM crops, organic agriculture and breeding for sustainability. Sustainability 6:4273–4286
Chae L, Kim T, Nilo-Poyanco Rhee SY (2014) Genomic signatures of specialized metabolism in plants. Science 244:510–513
Chalk PM (2016) The strategic role of ^{15}N in quantifying the contribution of endophytic N_2 fixation to the nutrition of non-legumes. Symbiosis 69:63–80
Chen Q, Liu Z, Wang B, Wang X, Lai J, Tian F (2015) Transcriptome sequencing reveals the roles of transcription factors in modulating genotype by nitrogen interaction in maize. Plant Cell Rep 34:1761–1771
Ciampitti IA, Vyn TJ (2013) Grain nitrogen source changes over time in maize: a review. Crop Sci 53:366–377
Colmsee C, Masher M, Czauderna T, Hartman A, Schlüter U, Zellerhoff N, Schmitz J, Bräutigam A, Pick TR, Alter P, Gahrtz M, Witt S, Fernie AR, Börnke F, Fahenenstich H, Bucher M, Dresselahaus T, Weber APM, Schreiber F, Scholtz U, Sonnewald U (2012) OPTIMAS-DW: a comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize. BMC Plant Biol 12:245
Coque M, Gallais A (2007) Genetic variation among European maize varieties for nitrogen use efficiency under low and high nitrogen fertilization. Maydica 52:383–397
Cren M, Hirel B (1999) Glutamine synthetase in higher plants: regulation of gene and protein expression from the organ to the cell. Plant Cell Physiol 40:1187–1193
Davenport S, Lay PL, Sanchez-Tamburrino JP (2015) Nitrate metabolism in tobacco leaves overexpressing *Arabidopsis* nitrite reductase. Plant Physiol Biochem 97:96–107
De Abreu e Lima F, Westhues M, Cuardos-Inostroza A, Willmitzer L, Melchinger AE, Nikoloski F (2017) Metabolic robustness in young roots underpins a predictive model of maize hybrid performance in the field. Plant J 90:319–329
Dechorgnat J, Francis KL, Dhugga KS, Rafalski JA, Tyerman SD, Kaiser BN (2018) Root ideotype influences nitrogen transport and assimilation in maize. Front plant Sci 9:531
Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hliaing AL, Aung HH, Nelissen H, Baute J, Frascaroli E, Churchill GA, Inzé D, Morgante M, Pè ME (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome Biol 16:167
Deng M, Li D, Lou J, Xiao Y, Liu H, Pan Q, Zhang X, Jin M, Zhao M, Yan J (2017) The genetic architecture of amino acid dissection by association and linkage analysis in maize. Plant Biotech J 1–14
de Souza R, Ambrosini A, Passaglla LMP (2015) Plant growth-promoting bacteria as inoculants in agricultural soils. Genet Mol Biol 38:401–419
Dubois F, Tercé-Laforgue T, Gonzalez-Moro MB, Estavillo MB, Sangwan R, Gallais A, Hirel B (2003) Glutamate dehydrogenase in plants: is there a new story for an old enzyme? Plant Physiol Biochem 41:565–576
Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Allen F, Hurst R, White B, Kernodle SP, Bromley JR, Sanchez-Tamburrino JP, Lewis RS, Mueller LA (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. BMC Genomics 18:448
EEA (2012) Air quality in Europe, 2012-report, EAA report no. 4/2012. European Environment Agency, Copenhagen, Denmark, p 104. https://doi.org/10.28000/55823
Erdal S, Turk H (2016) Cysteine-induced upregulation of nitrogen metabolism-related genes and enzymes activities enhance tolerance of maize seedlings to cadmium stress. Environ Exp Bot 132:92–99
Espejo-Herrera N, Gràcia-Lavedan E, Boldo E et al (2016) Colorectal cancer risk and nitrate exposure through drinking water and diet. Int J Cancer 139:334–346
Fageria NK (2008) The use of nutrients in crop plants. CRC Press, Taylor & Francis group, Boca Raton, London, New York
Fan X, Tang Z, Tan Y, Zhang Y, Luo B, Yang M, Lian X, Shen Q, Miller AJ, Xu G (2016) Overexpression of a pH-sensitive nitrate transporter in rice increases crop yield. Proc Natl Acad Sci USA 113:7118–7123
Farré G, Twyman RM, Christou P, Capell T, Zhu C (2015) Knowledge-driven approaches for engineering complex metabolic pathways in plants. Curr Opin Biotech 32:54–60
Fernie AR, Stitt M (2012) On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and network interactions. Plant Physiol 158:1139–1145
Fischer JJ, Beatty PH, Good AG, Muench DG (2013) Manipulation of microRNA expression to improve nitrogen use efficiency. Plant Sci 210:70–81
Fixen P, Brentup F, Bruuslema TW, Garcia F, Norton R, Zingore S (2015). Nutrient/fertilizer use efficiency: measurement, current situation and trends. In: Drechel P, Heffer P, Magen H, Mikkelsen R, Wilchelns D (eds) Managing water and fertilizer for sustainable agricultural intensification. International Fertilizer Industry Association (IFA), International Water Management Institute (IWMI), International Plant Nutrition Institute (IPNI), and International Potash Institute (IPI), Paris, France, January 2015, p 8

Fontaine JX, Tercé-Laforgue T, Armengaud P, Clément G, Renou JP, Pelletier S, Catterou M, Azzopardi M, Gibon Y, Lea PJ, Hirel B, Dubois F (2012) Characterization of a NADH-dependent glutamate dehydrogenase mutant of Arabidopsis demonstrates the key role of this enzyme in root carbon and nitrogen metabolism. Plant Cell 24:4044–4065

Forde BG, Lea PJ (2007) Glutamate in plants: metabolism, regulation, and signalling. J Exp Bot 58:2339–2358

Fowler D, Steadman CE, Stevenson D, Coyle M, Ress RM, Skiba UM, Sutton MA, Cape JN, Dore AJ, Vieno M, Simpson D, Zaehle S, Stocker BD, Rinaldi M, Facchini MC, Flecahrd CR, Nemitz E, Twigg M, Erisman JW, Butterbach-Bahlk K, Galloway JN (2015) Effects of global change during the 21st century on the nitrogen cycle atmos. Chem Phys 15:13849–13893

Fox T, DeBruin J, Haug Collet K, Trimnel M, Clapp J, Leonard A, LI B, Scolaro E, Collinson S, Glassman K, Miller M, Schlusser J, Dolan D, Lu L, Gho C, Albertsen M, Loussaert D, Shen B (2017) A single point mutation in *Ms44* results in dominant male sterility and improves nitrogen use efficiency in maize. Plant Biotech J 15:942–952

Freund DM, Hegeman AD (2017) Recent advances in stable isotope-anabled mass spectrometry-based plant metabolomics. Curr Opin Plant Biol 43:41–48

Fukushima A, Kusano M (2013) Recent progress in the development of metabolome databases for plant systems biology. Front Plant Sci 4:73

Fukushima A, Kusano M (2014) A network perspective on nitrogen metabolism from model to crop plants using integrated 'omics' approaches. J Exp Bot 65:5619–5630

Gallais A, Coque M (2005) Genetic variation and selection for nitrogen use efficiency in maize: a synthesis. Maydica 50:531–547

Gallais A, Coque M, Quilleré I, Le Gouis J, Prioul JL, Hirel B (2007) Estimating proportions of N-remobilization and of post-silking N-uptake allocated to maize kernels by ^{15}N-labelling. Crop Sci 47:685–691

Gallais A, Hirel B (2004) An approach to the genetics of nitrogen use efficiency in maize. J Exp Bot 55:295–306

Gallavotti A, Whipple CJ (2015) Positional cloning in maize (*Zea mays* Subsp. *mays*, Poaceae). Appl Plant Sci 3:1400092

Galloway JN, Leach AM, Bleeker A, Erisman JW (2013) A chronology of human understanding of the nitrogen cycle. Phil Trans Royal Soc Lond B 368. https://doi.org/10.1098/rstb.2013.0120

Garnett T, Conn V, Plett D, Conn S, Zanghellini J, Mackenzie N, Enju A, Francis K, Holtham L, Roessner U, Boughton B, Bacic A, Shirley N, Rafalski A, Dhugga K, Tester M, Kaiser BN (2013) The response of the maize nitrate transport system to nitrogen demand and supply across the lifecycle. New Phytol 198:82–94

Garnett T, Plett D, Conn V, Conn S, Rabie H, Rafalski JA, Dhugga K, Tester MA, Kaiser B (2015) Variation for N uptake system in maize: genotypic response to N supply. Front plant Sci 6:936

Gehan MA, Greenham K, Mockler TC, McKlung CR (2015) Transcriptional networks-crops, clocks, and abiotic stress. Curr Opin Plant Biol 24:39–46

Gellings CW, Parmenter KE (2016) In: Gellings CW (ed) Efficient use and conservation of energy. Energy efficiency in fertilizers production and use, vol II. EOLSS Publications, p 123

Gu R, Duan F, An X, Zhang F, von Wirén N, Yuan L (2013) Characterization of AMT—mediated high affinity ammonium uptake in roots of maize (*Zea mays* L.). Plant Cell Physiol 54:1515–1514

Guo S, Sun WY, Gu RL, Zhao BQ, Yuan LX, Mi GH (2014) Expression of genes related to nitrogen metabolism in maize grown under organic and inorganic nitrogen supplies. Soil Sci Plant Nutr 61:275–280

Gupta R, Hou Z, Loussaert DL, Shen B, Wood LK (2013) Engineering plants for efficient uptake and utilization of urea to improve crop production. US Patent 20140351998 A1

Habbib H, Verzeaux J, Nivelle E, Roger D, Lacoux J, Catterou M, Hirel B, Dubois F, Tétu T (2016) Conversion to no-till improves maize nitrogen use efficiency in a continuous cover cropping system. PLOS One 11:10. https://doi.org/10.1371/journal.pone.0164234

Habbib H, Hirel B, Verzeaux J, Roger D, Lacoux J, Lea PJ, Dubois F, Tetu T (2017) Investigating the combined effect of tillage, nitrogen fertilization and cover crops on nitrogen use efficiency in winter wheat. Agronomy 7:66. https://doi.org/10.3390/agronomy704006

Habermeyer M, Roth A, Guth S, Diel P, Engel KH, Epe B, Fürst P, Heinz V, Humpf HU, Joost HG, Knorr D, de Kok T, Kulling S, Lampen A, Marko D, Rechkemmer G, Rietjens I, Stadler RH, Vieths S, Vogel R, Steinberg P, Eisenbrand G (2015) Nitrate and nitrite in the diet: how to assess their benefit and risk for human health. Mol Nutr Food Res 59:106–128

Haegele JW, Cook KA, Nichols DM, Below FE (2013) Changes in nitrogen use traits associated with genetic improvement for grain yield of maize hybrids released in different decades. Crop Sci 53:1256–1268

Han J, Wang L, Zheng H, Pan X, Li H, Chen F, Li X (2015a) ZD958 is a low-nitrogen-efficient maize hybrid at the seedling stage among five maize and two teosinte lines. Planta 242:935–949

Han M, Okamoto M, Beatty PH, Rothstein SJ, Good AJ (2015b) The genetics of nitrogen use efficiency in crop plants. Ann Rev Genet 49:269–289

He C-M, Liu C-X, Liu Q, Gao X-X, Li N, Zhang J-R, Wang L-M Liu T-S (2014) Over-expression of glutamine synthetase genes *Gln1-3/Gln1-4* improved nitrogen assimilation and maize yields. Maydica 59:250–256

Hirel B, Bertin P, Quillere I, Bourdoncle W, Attagnant C, Dellay C, Gouy A, Cadiou S, Retailliau C, Falque M, Gallais A (2001) Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiol 125:1258–1270

Hirel B, Gallais A (2011) Nitrogen use efficiency—physiological, molecular and genetic investigations towards crop improvement. In: Prioul JL, Thévenot C, Molnar T (eds) Advances in maize, essential reviews in experimental biology, vol 3. Society for Experimental Biology, UK, pp 285–310

Hirel B, Lea PJ (2001) Ammonia assimilation. In: Lea PJ, Morot-Gaudry JF (eds) Plant nitrogen. Springer, INRA Editions, pp 79–99

Hirel B, Lea PJ (2002) The biochemistry, molecular biology, and genetic manipulation of primary ammonia assimilation. In: Foyer CH, Noctor G (eds) Photosynthetic nitrogen assimilation and associated carbon and respiratory metabolism, advances in photosynthesis and respiration. Kluwer Academic Publishers, Dordrecht, Boston, London, pp 71–92

Hirel B, Le Gouis J, Bernard M, Perez P, Falque M, Quétier F, Joets J, Montalent P, Rogwoski P, Murigneux A, Charcosset A (2007a) Genomics and plant breeding: maize and wheat. In: Morot-Gaudry JF, Lea PJ, Briat JF (eds) Functional plant genomics. Science Publishers, Enfield, NH, pp 614–635

Hirel B, Le Gouis J, Ney B, Gallais A (2007b) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58:2369–2387

Hirel B, Martin A, Tercé-Laforgue T, Gonzalez-Moro MB, Estavillo JM (2005) Physiology of maize I: a comprehensive and integrated view of nitrogen metabolism in a C_4 plant. Physiol Plant 124:167–177

Hirel B, Perez P (2016) Increased kernel productivity of plants through the modulation of glutamine synthetase activity. US patent 9422571 B2

Hirel B, Tétu T, Lea PJ, Dubois F (2011) Improving nitrogen use efficiency in crops for a sustainable agriculture. Sustainability 3:1452–1485

Hodge A, Storer K (2015) Arbuscular mycorrhiza and nitrogen: implications for individual plants through to ecosystem. Plant Soil 386:1–19

Horde NG, Conley MN (2017) Regulation of dietary nitrate and nitrite: balancing essential physiological roles with potential health risks. In: Bryan NS, Loscalzo J (eds) Nitrite and nitrate in human health and disease, nutrition and health. Springer, Heidelberg, p 153

Hu J, Rampitsch C, Bykova NV (2015) Advances in plant proteomics toward improvement of crop productivity and stress resistance. Front Plant Sci 6:209

Humbert S, Subedi S, Cohn J, Zeng B, Bi YM, Chen X, Zhu T, McNicholas PD, Rothstein SJ (2013) Genome-wide expression profiling of maize in response to individual and combined water and nitrogen stresses. BMC Genomics 14:3

Ibrahim A, JIn X-L, Zhang YB, Cruz J, Vichyavichen P, Esiobu N, Zhang XH (2017) Tobacco plants expressing the maize nitrate transporter ZmNrt2.1 exhibit altered response of growth and gene expression to nitrate and calcium. Bot Stud 58:51

Jansen C, Zhang Y, Liu H, Gonzalez-Portilla PJ, lauter N, Kumar B, Trucillo-Silva I, San Martin JP, Lee M, Simcox K, Schussler J, Dhugga K, Lübbersedt (2015) Genetic and agronomic assessment of cob traits in corn under low and normal nitrogen management conditions. Theor Appl Genet 128:1231–1242

Jin X, Li W, Hu D, Shi X, Zhang X, Zhang F, Fu Z, Ding D, Liu Z, Tang J (2015) Biological responses and proteomic changes in maize seedlings under nitrogen deficiency. Plant Mol Biol Rep 33:490–504

Junker BH (2014) Flux analysis in plant metabolic networks: increasing throughput and coverage. Curr Opin Plant Biol 26:18188

Kanter DR, Zhang X, Mauzerall DL (2015) Reducing nitrogen pollution while decreasing farmer's costs and increasing fertilizer industry profits. J Environ Qual 44:325–335

Klopfenstein TJ, Erickson GE, Berger LL (2013) Maize is critically important source of food, feed, energy and forage in the USA. Field Crop Res 153:5–11

Kou HP, Li Y, Song XX, Ou XF, Xing SC, Ma J, Von Wettstein D (2011) Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (*Oryza sativa* L.). J Plant Physiol 168:1685–1693

Kruger NJ, Ratcliffe RG (2012) Pathways and fluxes: exploring the plant metabolic network. J Exp Bot 63:2243–2246

Kruger NJ, Ratcliffe RG (2015) Fluxes through metabolic network: measurements, predictions insights and challenges. Biochem J 465:27–38

Kuan KB, Othman R, Abdul Rahim K, Shamsuddin ZH (2016) Plant growth promoting rhizobacteria inoculation to enhance vegetative growth, nitrogen fixation and nitrogen remobilization of maize under greenhouse conditions. PLOS One. https://doi.org/10.1371/journal.pone.0152478

Labboun S, Tercé-Laforgue T, Roscher A, Bedu M, Restivo FM, Velanis CN, Skopelitis DS, Moshou PN, Roubelakis-Angelakis KA, Suzuki A, Hirel B (2009) Resolving the role of plant glutamate dehydrogenase: I. *In vivo* real time nuclear magnetic resonance spectroscopy experiments. Plant Cell Physiol 50:1761–1773

Lassaletta L, Billen G, Garnier J, Bouwman L, Velazquez E, Mueller ND, Gerber JS (2016) Nitrogen use in the global food system: past trends and future trajectories of agronomic performance, pollution, trade, and dietary demand. Environ Res Lett 11:095007

Lea PJ, Miflin BJ (2011) Nitrogen assimilation and its relevance to crop improvement. In: Foyer CH, Zhang H (eds) Annual plant reviews, nitrogen

metabolism in plants in the post-genomic era. Wiley-Blackwell, Chichester, UK, vol 42, p 40

Lea PJ, Sodek L, Parry MAJ, Shewry PR, Halford NG (2007) Asparagine in plants. Ann Appl Biol 150:1–26

Léran S, Varala K, Boyer JC, Chiurazzi M, Crawford N, Daniel-Vedele F, David L, Dickstein R, Fernandez E, Forde B, Gassmann W, Geiger D, Gojon A, Gong JM, Halkier BA, Harris JM, Hedrich R, Limami AM, Rentsh D, Seo M, Tsay LF, Zhang M, Coruzzi G, Lacombe B (2014) A unified nomenclature of nitrate TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. Trends Plant Sci 19:5–9

Li MG, Villemur R, Hussey PJ, Silflow CD, Gantt JS, DP (1993) Differential expression of six glutamine synthetase genes in *Zea mays*. Plant Mol Biol 23:401–40

Li Y, Wang M, Zhang F, Xu Y, Chen X, Qin X, Wen X (2016) Effect of post-silking drought on nitrogen partitioning and gene expression pattern of glutamine synthetase and asparagine synthetase in two maize (*Zea mays* L.) varieties. Plant Physiol Biochem 102:62–69

Liao C, Peng Y, Ma W, Liu R, LI C, Li X (2012) Proteomic analysis revealed nitrogen-mediated metabolic, developmental, and hormonal regulation of maize (*Zea mays* L.) ear growth. J Exp Bot 63:5275–5288

Lightfoot DA, Mungur R, Ameziane R, Nolte S, Long L, Bernhard K, Colter A, Jones K, Iqbal MJ, Varsa E, Young B (2007) Improved drought tolerance of transgenic *Zea mays* plants that express the glutamate dehydrogenase gene (*gdhA*) of *E. coli*. Euphytica 156:103–116

Limami AM, Rouillon C, Glevarec G, Gallais A, Hirel B (2002) Genetic and physiological analysis of germination efficiency in maize in relation to N metabolism reveals the importance of cytosolic glutamine synthetase. Plant Physiol 130:1860–1870

Lisec J, Römish-Margi L, Nikoloski Z, Piepho HP, Giavalisco P, Selbig J, Gierl A, Willmitzer L (2011) Corn hybrids display lower metabolite variability and complex inheritance patterns. Plant J 66:326–336

Liu GW, Sun AL, Li DQ, Athman A, Giliham M, Liu LH (2015a) Molecular identification and functional analysis of a maize (*Zea mays*) DUR3 homolog that transports urea with high affinity. Planta 241:861–874

Liu H, Niu H, Gonzalez-Portilla P, Zhou Z, Wang L, Zuo T, Qin C, Tai S, Jansen C, Shen Y, Lin H, Lee M, Ware D, Zhang Z, Lübberstedt T, Pan G (2015b) An ultra-high-density map as a community resource for dissecting the genetic basis of quantitative traits in maize. BMC Genomics 16:1078

Lohaus G, Büker M, Hubman M, Soave C, Heldt H (1998) Transport of amino acids with special emphasis on the synthesis and transport in the Illinois low protein and Illinois high protein strains of maize. Planta 205:181–188

Luo B, Tang H, Liu H, Shunzong S, Zhang S, Wu L, Liu D, Gao S (2015) Mining for low-nitrogen tolerance genes by integrating meta-analysis and large-scale gene expression data from maize. Euphytica 206:117–131

Luo J (2015) Metabolite-based genome—wide association studies in plants. Curr Opin Plant Biol 24:31–38

Lupini A, Meracti F, Araniti F, Miller AJ, Sunseri F, Abenavoli MR (2016) NAR2.1/NRT2.1 functional interaction with NO_3^- and H^+ fluxes in high-affinity nitrate transport in maize roots. Plant Physiol Biochem 102:107–114

Lupini A, Araniti F, Mauceri A, Princi MP, Sorgonà A, Sunseri F, Varanini Z, Abenavoli MR (2018) Coumarin enhances nitrate uptake in maize roots through modulation of plasma membrane H^+ATPase activity. Plant Biol 20:390–398

Lv Y, Liang Z, Ge M, Qi W, Zhang T, Lin F, Peng Z, Zhao H (2016) Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). BMC Genomics 17:350

Martin A, Lee J, Kichey T, Gerentes D, Zivy M, Tatou C, Balliau T, Valot B, Davanture M, Dubois F, Tercé-Laforgue T, Coque M, Gallais A, Gonzalez-Moro MB, Bethencourt L, Quilleré I, Habash DZ, Lea PJ, Charcosset A, Perez P, Murigneux A, Sakakibara H, Edwards KJ, Hirel B (2006) Two cytosolic glutamine synthetase isoforms of maize (*Zea mays* L.) are specifically involved in the control of grain production. Plant Cell 18:3252–3274

McAllister CH, Beatty PH, Good AG (2012) Engineering nitrogen use efficient crop plants: the current status. Plant Biotech J 10:1011–1025

McKenzie FC, Williams J (2015) Sustainable food production: constraints, challenges and choices by 2050. Food Secur 7:221–223

McNally SF, Hirel B, Gadal P, Mann AF, Stewart GR (1983) Glutamine synthetase in higher plants. Evidence for a specific isoform content related to their possible physiological role and their compartmentation within the leaf. Plant Physiol 72:22–25

Mengel K, Robin P, Salsac L (1983) Nitrate reductase activity in shoots and roots of maize seedlings as affected by the form of nitrogen nutrition and the pH of the solution. Plant Physiol 71:618–622

Miflin BJ, Habash DZ (2002) The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. J Exp Bot 53:979–987

Moss B (2008) Water pollution in agriculture. Phil Trans Royal Soc Lond B 363:659–666

Muhitch MJ (2003) Distribution of the glutamine synthetase isozyme GS_{p1} in maize (*Zea mays*). J Plant Physiol 160:601–605

Nakabayashi R, Saito K (2015) Integrated metabolomics for abiotic stress response in plants. Curr Opin Plant Biol 24:10–16

Nazir M, Pandey R, Siddiqi TO, Ibrahim MM, Qureshi MI, Abraham G, Vengavasi K, Ahmad A

(2016) Nitrogen-deficiency stress induces protein expression differentially in low-N tolerant and low-N sensitive maize genotypes. Front Plant Sci 7:298

Nieder R, Benbi DK, Scherer HW (2011) Fixation and defixation of ammonium in soils: a review. Biol Fertil Soils 47:1–14

Nouri MZ, Ghaffari MR, Sobhanain H, Hajirezaei MR (2016) Proteomics approach for identification of nutrient deficiency related proteins in crop plants. In: Salekdeh GH (ed) Agricultural proteomics, vol 2. Springer, pp 177–201

Oaks A (1994) Efficiency of nitrogen utilization in C_3 and C_4 cereals. Plant Physiol 106:407–414

Obata T, Witt S, Lisec J, Palacios-Rojas N, Florez-Saraza I, Yousfi S, Araus JL, Cairns JE, Fernie AR (2015) Metabolite profile of maize leaves in drought, heat, and combined stress field trials reveal the relationship between metabolism and grain yield. Plant Physiol 169:2665–2683

O'Brien JA, Vega A, Bouguyon E, Krouk G, Gojon A, Coruzzi G, Guttiérez RA (2016) Nitrate transport, sensing, and response in plants. Mol Plant 9:837–856

Oita A, Malik A, Kanemoto K, Geschke A, Nishijima S, Lenzen M (2016) Substantial nitrogen pollution embedded in international trade. Nat Geosc 9:111–115

Pan X, Hasan MM, Li Y, Liao C, Zheng H, Liu R, Li X (2015) Asymmetric transcriptomic signatures in the maize era under optimal- and low-nitrogen conditions at silking, and functional characterization of amino acid transporters ZmAAP4 and ZmVAAT3. J Exp Bot 66:6149–6166

Parnell JJ, Berka R, Young HA, Sturino JM, Kang Y, Barnhart DM, DiLeo MV (2016) From the lab to the farm: an industrial perspective of plant beneficial microorganisms. Front Plant Sci 7:1110

Pathak RR, Lochab S, Raghuram N (2011) Improving nitrogen-use efficiency. Comprehensive biotechnology, vol 4, pp 209–218

Peña PA, Quach T, Sato S, Ge Z, Nersesian N, Changa T, Dweikat I, Soundararajan M, Clemente TE (2017) Expression of the maize *Dof1* transcription factor in wheat and sorghum. Front Plant Sci 8:434

Plénet D, Lemaire G (1999) Relationships between dynamics of nitrogen uptake and dry matter accumulation in maize crops. Determination of critical N concentration. Plant Soil 216:65–82

Plett D, Bauman U, Schreiber A, Holtham L, Kalashyan E, Toubia J, Nau J, Beatty M, Rafalski A, Dhugga KS, Tester M, Garnett T, Kaiser BN (2015) Maize maintains growth in response to decreased nitrate supply through a highly dynamic and developmental stage-specific transcriptional response. Plant Biotech J 14:342–353

Plett D, Holtham L, Bauman U, Kalashyan E, Francis K, Enju A, Toubia J, Roessner U, Bacic A, Rafalski A, Dhugga KS, Tester M, Garnett T, Kaiser BN (2016) Nitrogen assimilation system in maize is regulated by developmental and tissue-specific mechanisms. Plant Mol Biol 92:293–312

Poorter H, Anten NSP, Marcelis LFM (2013) Physiological mechanisms in plant growth models: do we need a supra-cellular systems biology approach? Plant Cell Environ 36:1673–1690

Prinsi B, Espen L (2015) Mineral nitrogen source differently affect root glutamine synthetase isoforms and amino acid balance among organs in maize. BMC Plant Biol 15:96

Raihan MS, Liu J, Huang J, Guo H, Pan Q, Yan J (2016) Multi-environment QTL analysis of grain morphology traits and fine mapping of a kernel-width QTL in Zheng58 x SK maize population. Theor Appl Genet 129:1465–1477

Reetz HF, Heffer P, Bruulsema TW (2015) 4R nutrient stewardship: a global framework for sustainable fertilizer management. In: Drechsel P, Heffer P, Magen H, Mikkelsen R, Wilchelns (eds) Managing water and fertilizer for sustainable agricultural intensification. International Fertilizer Industry Association (IFA), International Water Management Institute (IWMI), International Plant Nutrition Institute (IPNI), and International Potash Institute (IPI), Paris, France, Jan 2015, p 65

Ren B, Dong S, Zhao B, Liu P, Zhang J (2017) Responses of nitrogen metabolism, uptake and translocation of maize to waterlogging at different growth stages. Front Plant Sci 8:1216

Richardson AE, Barea JM, McNeill AM, Prigent-Combaret C (2009) Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. Plant Soil 321:305–339

Riedelsheimer C, Czedick-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012a) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genet 44:217–222

Riedelsheimer C, Lisec J, Czedick-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012b) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci USA 109:8872–8887

Roach E, Duiker SW, Chopra S (2016) Soil management affects expression of genes involved in carbon and nitrogen metabolism in maize. Crop Sci 56:1841–1856

Sakakibara H, Kawabata S, Takahashi H, Hase T, Sugiyama T (1992) Molecular cloning of the family of glutamine synthetase genes from maize: expression of genes for glutamine synthetase and ferredoxin-dependent glutamate synthase in photosynthetic and non-photosynthetic tissues. Plant Cell Physiol 33:49–58

Salvi S, Tuberosa R (2007) Cloning QTLs in Plants. In: Varshney RK, Tuberosa R (eds) Genomics-Assisted Crop Improvement. Springer, Dordrecht, The Netherlands, vol 1, pp 207–226

Sangwan RS, Bourgeois Y, Brown S, Vasseur G. Sangwan-Norreel B (1992) Characterization of

competent cells and early events of *Agrobacterium*-mediated genetic transformation in *Arabidopsis thaliana*. Planta 188: 439–456

Schilmiller AL, Pichersky E, Last RL (2012) Taming the hydra of specialized metabolism: how systems biology and comparative approaches are revolutionizing plant biochemistry. Cur Opin Plant Biol 15:338–344

Schlüter U, Colmsee C, Scholtz U, Bräutigam A, Weber APM, Zellerhoff N, Bucher M, Fahnenstich H, Sonnewald U (2013) Adaptation of maize source leaf metabolism to stress related disturbances in carbon, nitrogen and phosphorus balance. BMC Genomics 14:442

Schmidt RR, Miller P (1999) Polypeptides and polynucleotides relating to the and subunits of a glutamate dehydrogenase and methods of use. United States Patent no 5, 879, 941, Mar 9

Seebauer JR, Moose SP, Fabbri BJ, Crossland LD, Below FE (2004) Amino acid metabolism in maize earshoots. Implications for assimilate preconditioning and nitrogen signaling. Plant Physiol 136:4326–4334

Seebauer JR, Singletary GW, Krumpelman PM, Ruffo ML, Below FE (2010) Relationship of source and sink in determining kernel composition in maize. J Exp Bot 61:551–519

Shachar-Hill Y (2013) Metabolic network flux analysis for engineering plant systems. Curr Opin Plant Biol 24:247–255

Shen M, Broeckling CD, Chu EY, Ziegler G, Baxter IR, Prenni JE, Hoekenga OA (2013) Leveraging non-targeted metabolite profiling via statistical genomics. PLOS One 8:e5767

Simons M, Saha R, Amiour A, Kumar A, Guillard L, Clément G, Miquel M, Li Z, Mouille G, Lea PJ, Hirel B, Maranas CD (2014a) Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. Plant Physiol 166:1659–1674

Simons M, Saha R, Guillard L, Clément G, Armengaud P, Cañas R, Maranas CD, Lea PJ, Hirel B (2014b) Nitrogen use efficiency in maize (*Zea mays* L.): from "omics" studies to metabolic modelling. J Exp Bot 65:5657–5671

Singh R, Pankaj Sahu P, Muthamilarasan M, Dhaka A, Prasad M (2017) Genomics-assisted breeding for improving stress tolerance of graminaceous crops to biotic and abiotic stresses: progress and prospects. In: Senthil-Kumar M (ed) Plant tolerance to individual and concurrent stresses. Springer, New Delhi, India, pp 59–81

Singh RP, Srivastava HS (1986) Increase in glutamate synthase (NADH) activity in maize seedlings in response to nitrate and ammonium nitrogen. Physiol Plant 66:413–416

Singletary GW, Doehlert DC, Wilson CM, Muhitch MJ, Below FE (1990) Response of enzymes and storage proteins of maize endosperm to nitrogen supply. Plant Physiol 94:858–864

Sinha SK, Srinivasan R, Mandal PK (2015) MicroRNA-based approach to improve nitrogen use efficiency in crop plants. In: Rakshit A, Sigh AB, Sen A (eds) Nutrient use efficiency: from basic to advances, Springer, India, pp 221–235

Sirohi G, Pandey B, Deveshwar P, Giri J (2016) Emerging trends in epigenetic regulation of nutrient deficiency response in plants. Mol Biotech 58:159–171

Skopelitis DS, Paranychiankis NV, Paschalidis KA, Plianokis ED, Delis ID, Yakoumakis DI, Kouvarakis A, Papadakis ED, Stephanou EG, Roubelakis-Angelakis KA (2006) Abiotic stress generates ROS that signal expression of anionic glutamate dehydrogenase to form glutamate for proline synthesis in tobacco and grapevine. Plant Cell 18:2767–2781

Skopelitis DS, Paranychiankis NV, Kouvarakis A, Spyros A, Stephanou EG, Roubelakis-Angelakis KA (2007) The isoenzyme 7 of tobacco NADH-dependent glutamate dehydrogenase exhibits high deaminating and low aminating activity. Plant Physiol 145:1–9

Smith KA (2017) Changing views of nitrous oxide emissions from agricultural soil: key controlling processes and assessment at different spatial scales. Eur J Soil Sci 68:137–155

Song W, Wang B, Hauck AL, Dong X, Li J, Lai J (2016) Genetic dissection of maize seedling root system architecture traits using an ultra-high-density bin-map and recombinant inbred line population. J Int Plant Biol 58:266–279

Sorgonà A, Lupini A, Mercati F, Di Dio L, Sunseri F, Abenavoli MR (2011) Nitrate uptake along the maize primary root: an integrated physiological and molecular approach. Plant Cell Envir 34:1127–1140

Srivastava A, Kowalski GM, Callahan DL, Meikle PJ, Creek DJ (2016) Strategies for extending metabolomics studies with stable isotope labeling and fluxomics. Metabolites 6:32

Su C, Wang W, Gong S, Zuo J, Li S, Xu S (2017) High density linkage map construction and mapping of yield DTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. Front Plant Sci 8:706

Sun CX, Li MQ, Gao XX, Liu LN, WU XF, Zhou JH (2016) Metabolic response of maize plants to multi-factorial abiotic stresses. Plant Biol 18:120–129

Swain EY, Rempelos L, Orr CH, Hall G, Chapman R, Almadni M, Stockdale EA, Kidd J, Leifert C, Cooper JM (2014) Optimizing nitrogen use efficiency in wheat and potatoes: interaction between genotypes and agronomic practices. Euphytica 199:119–136

Sweetlove LJ, Nielsen J, Fernie AR (2017) Engineering central metabolism—a grand challenge for plant biologists. Plant J 90:749–763

Tabuchi M, Abiko T, Yamaya T (2007) Assimilation of ammonium ions and reutilization of nitrogen in rice (*Oryza sativa* L.). J Exp Bot 58:2319–2327

Tercé-Laforgue T, Dubois F, Ferrario-Mery S, Pou de Crecenzo MA, Sangwan R, Hirel B (2004) Glutamate dehydrogenase of tobacco (*Nicotiana tabacum* L.) is mainly induced in the cytosol of phloem companion

cells when ammonia is provided either externally or released during photorespiration. Plant Physiol 136:4308–4317

Tercé-Laforgue T, Bedu M, Dargel-Graffin C, Dubois F, Gibon Y, Restivo FM, Hirel B (2013) Resolving the role of plant glutamate dehydrogenase: II. Physiological characterization of plants overexpressing individually or simultaneously the two enzyme subunits. Plant Cell Physiol 54:1634–1647

Tercé-Laforgue T, Clément G, Marchi L, Restivo FM, Lea PJ, Hirel B (2015) Resolving the role of plant NAD-glutamate dehydrogenase: III. Overexpressing individually or simultaneously the two enzyme subunits under salt stress induces changes in the leaf metabolic profile and increases plant biomass production. Plant Cell Physiol. 56:1918–1929

Thomsen HC, Erikson D, MØller IS, Schjoerring JK (2014) Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency? Trends Plant Sci 19:656–663

Todd J, Screen S, Crowley J, Peng J, Andersen S, Brown T, Qi Q, Fabbri B, Duff SMG (2008) Identification and characterization of four distinct asparagine synthetase (*AsnS*) genes in maize (*Zea mays* L.). Plant Sci 175:799–808

Tohge T, Scossa F, Fernie AR (2015) Integrative approaches to enhance understanding of plant metabolic pathway structure and regulation. Plant Physiol 169:1499–1511

Toubiana D, Xue W, Zhang N, Kremling K, Gur A, Pilosof S, Gibon Y Stitt M, Buckler ES, Fernie AR, Fait A (2016) Correlation based network analysis of metabolite and enzyme profiles reveals a role of citrate biosynthesis in modulating N and C metabolism in *Zea mays*. Front Plant Sci 7:1022

Trevisan S, Nonis A, Begheldo M, Manoli A, Palme K, Caporale G, Ruperti B, Quaggiotti S (2012) Expression and tissue-specific localization of nitrate-responsive miRNAs in roots of maize seedlings. Plant Cell Environ 35:1137–1155

Trevisan S, Manoli A, Ravazzolo L, Botton A, Pivato M, Masi A, Quagiotti S (2015) Nitrate sensing by the maize root apex transition zone: a merged transcriptomic and proteomic survey. J Exp Bot 66:3699–3715

Trucillo Silva I, Abbaraju HKR, Fallis LP, Liu H, Lee M, Dhugga KS (2017) Biochemical and genetic analyses of N metabolism in maize testcross seedlings. Theor Appl Genet 130:1453–1466

Urano K, Kurihara Y, Seki M, Shinozaki K (2010) 'Omics' analysis of regulatory network in plant abiotic stresses. Curr Opin Plant Biol 13:132–138

Uribelarrea M, Below FE, Moose SP (2004) Kernel composition and productivity of maize hybrids derived from the Illinois protein strains in response to variable nitrogen supply. Crop Sci 44:1593–1600

Valadier MH, Yoshida A, Grandjean O, Morin H, Kronenberger J, Boutet S, Raballand A, Hase T, Yoneyama T, Suzuki A (2008) Implication of the glutamine synthetase/glutamate synthase pathway in conditioning the amino acid metabolism in bundle sheath and mesophyll cells of maize. FEBS J 275:3193–3206

Verzeaux J, Alahmad A, Habbib H, Nivelle E, Roger D, Lacoux J, Decocq G, Hirel B, Catterou M, Spicher F, Dubois F, Duclercq J, TétuT (2016a) Cover crops prevent from deleterious effect of nitrogen fertilization on bacterial diversity by maintaining carbon concentration in plowed soil. Geoderma 281:49–57

Verzeaux J, Roger D, Lacoux J, Nivelle E, Adam C, Habbib H, Hirel B, Dubois F, TétuT (2016b) In winter wheat, no-till increases mycorrhizal colonization thus reducing the need for nitrogen fertilization. Agronomy 6:38

Verzeaux J, Hirel B, Dubois F, Lea PJ, TétuT (2017) Agricultural practices to improve nitrogen use efficiency through the use of arbuscular mycorrhizae: basic and agronomic aspects. Plant Sci 264:48–56

Wallace JG, Larsson SJ, Buckler ES (2014) Entering the second century of maize quantitative genetics. Heredity 112:30–38

Wang H, Loussaert DL (2015) Functional expression of yeast nitrate transporter (YNT1) and a nitrate reductase in maize. US patent 8975474 B2

Wei S, Wang X, Shi D, Li Y, Zhang J, Liu P, Zhao B, Dong S (2016) The mechanisms of low nitrogen induced weakened photosynthesis in summer maize (*Zea mays* L.) under field conditions. Plant Physiol Biochem 105:118–128

Wen W, Li K, Alseek S, Omranian N, Zhao L, Zhou Y, Xiao Y, Jin M, Yang N, Liu Florian A, Li W, Pan Q, Nikoloski Z, Yan J, Fernie AR (2015) Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. Plant Cell 27:1839–1856

Wen Z, Tyerman SD, Dechorgnat J, Ovchinnikova E, Dhugga KS, Kaiser BN (2017) Maize NFP6 proteins are homologs of Arabidopsis CHL1 that are selective for both nitrate and chloride. Plant Cell. https://doi.org/10.1105/tpc.16.00724

Widiez T, El Kafafi ES, Girin T, Berr A, Ruffel S, Krouk G, Vayssieres A, Shen WH, Coruzzi GM, Gojon A, Lepetit M (2011) High nitrogen insensitive 9 (HNI9)-mediated systemic expression of root NO_3^- uptake is associated with changes in histone methylation. Proc Natl Acad Sci USA 108:13329–13334

Withers PJA, Neal C, Jarvie HP, Doody DG (2014) Agriculture and eutrophication: where do we go from here? Sustainability 6:5853–5875

Wuyts N, Dhondt S, Inzé D (2015) Measurment of plant growth in view of an integrative analysis of regulatory network. Curr Opin Plant Biol 25:90–97

Xu Z, Zhong S, Li X, Li W, Rothstein SJ, Zhang S, Bi YM, Xie C (2011) Genome-wide identification of microRNAs in response to low nitrate availability in maize leaves and roots. PLOS One 6:e28009

Xu C, Ren Y, Jian Y, Guo Z, Zhang Y, Xie C, Fu J, Wang H, Wang G, Xu Y, Li P, Zou C (2017)

Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. Mol Breed 37:20

Yanagisawa S (2004) Dof domain proteins: plant–specific transcription factors associated with diverse phenomena unique to plants. Plant Cell Physiol 45:386–391

Yanagisawa S, Akiyama A, Kisaka H, Uschimiya H, Miwa T (2004) Metabolic engineering with Dof1 transcription factor in plants: improved nitrogen assimilation and growth under low-nitrogen conditions. Proc Natl Acad Sci USA 101:7833–7838

Yamaya T, Kusano M (2014) Evidence supporting distinct functions of three cytosolic glutamine synthetases and two NADH-glutamate synthases. J Exp Bot 19:5519–5525

Yan S, Du X, Wu F, Li L, Li C, Meng Z (2014) Proteomics insights into the basis of interspecific facilitation for maize (*Zea mays*) in faba bean (V*icia faba*)/maize intercropping. J Proteom 109:111–124

Yang S, Wu J, Ziegler TE, Yang X, Zayed A, Rajani MS, Zhou D, Basra A, Schachtman D, Peng M, Armstrong CL, Caldo RA, Morrell JA, Lacy M, Staub JM (2011) Gene expression biomarkers provide sensitive indicators of *in planta* nitrogen status in maize. Plant Physiol 157:1841–1852

Yesbergenova-Cuny Z, Dinant S, Martin-Magniette ML, Quillere I, Armengaud P, Monfalet P, Lea PJ, Hirel B (2016) Genetic variability of the phloem sap metabolite content of maize (*Zea mays* L.) during the kernel-filling period. Plant Sci 252:347–357

York LM, Silberbush M, Lynch JP (2016) Spatiotemporal variation of nitrate uptake kinetics within the maize (*Zea mays* L.) root system is associated with greater nitrate uptake and interactions with architectural phenes. J Exp Bot 67:3763–3775

Yu P, White P, Hochlodinger F, Li C (2014) Phenotypic plasticity of the maize root system in response to heterogeneous nitrogen availability. Planta 240:667–768

Yuan L, Loqué D, Kojima S, Rauch S, Ishimaya K, Inoue E, Takahashi H, von Wirén N (2007) The organization of high-affinity ammonium uptake in *Arabidopsis* roots depends on the spatial arrangement and biochemical properties of AMT1-type transporters. Plant Cell 19:2636–2652

Yuan S, Li Z, Li D, Yuan N, Hu Q, Luo H (2015) Constitutive overexpression of rice *MicroRNA528* alters plant development and enhances tolerance to salinity stress and nitrogen starvation in creeping bentgrass. Plant Physiol 169:576–593

Zamboni A, Astlfi S, Zuchi S, Pii Y, Guradini K, Tonomi P, Varanini Z (2014) Nitrate induction triggers different transcriptional changes in a high and low nitrogen use efficiency maize inbred line. J Integr Plant Biol 56:1080–1094

Zanin L, Zamboni A, Monte R, Tomasi N, Varanini Z, Cesco S, Pinton R (2015) Transcriptomic analysis highlights the reciprocal interactions of urea and nitrate for nitrogen acquisition by maize roots. Plant cell Physiol 56:532–548

Zeng DD, Qin R, Li M, Alamin Md, Jin KL, Liu Y, Shi CH (2017) The ferredoxin-dependent glutamate synthase (OsFd-GOGAT) participates in leaf senescence and nitrogen remobilization in rice. Mol Genet Genom 292:385–395

Zhang N, Gibon Y, Wallace JG, Lepak N, Li P, Dedow L, Chen C, So YS, Kremling K, Bradbury PJ, Brutnell T, Stitt M, Buckler ES (2015) Genome-wide association of carbon and nitrogen in the maize nested association mapping population. Plant Physiol 168:575–583

Zhao M, Tai H, Sun S, Zhang F, Xu Y, Li WX (2012) Cloning and characterization of maize miRNAs involved in responses to nitrogen deficiency. PLOS One 7:e29669

Zhou X, Lin J, Zhou Y, yang Y, Liu H, Zhang C, Tang D, Zhao X, Zhu Y, Liu X (2015) Overexpressing a fungal *CeGDH* gene improves nitrogen utilization and growth in rice. Crop Sci 55:811–820

Genomics of Cold Tolerance in Maize 17

Elisabetta Frascaroli and Pedro Revilla

Abstract

Maize originated in tropical areas and improving cold tolerance is an important breeding objective for cultivation in high latitudes. We review the main limitations in understanding and improving cold tolerance in maize and the contribution of genomics in dissecting the genetic basis of the trait and selecting better genotypes. Physiological analyses revealed that non-optimal temperature exerts detrimental effects on a multitude of metabolic functions at different growing stages, each under the control of independent gene sets. Loci controlling cold tolerance at different growing stages have been investigated by means of linkage mapping or genome-wide association, revealing that no major genes are responsible for the trait. This finding was confirmed in transcriptomic studies that always revealed multiple candidates, and a large amount of data is being collected that altogether will make it possible to obtain a more coherent picture of response to cold. To harness the increasing body of information available from the maize genome sequence and gene expression data, new bioinformatics tools will be helpful for integrating the big-data obtained from the large-scale genomics and phenomics experiments. With the enhancement of knowledge, plant science is shifting its focus from "explanatory" to "predictive" and from a plant breeding perspective the focus will be predicting the breeding value of the best genotypes by using molecular information. The future strategies for selection of cold tolerance will involve intensive genotyping, high-precision phenotyping and advanced statistical analyses to predict the optimal genotypes for more time- and cost-efficient breeding strategies.

17.1 Introduction

Maize (*Zea mays* L.) originated in tropical areas and reached the high latitudes in America prior to the arrival of Columbus. Five centuries ago, maize was introduced in the Old World and was successfully adapted to a remarkable variety of climates. The large adaptation ability has allowed maize to become one of the major crops in the world both for food and feed and, more recently, as a source of renewable energy. Maize-growing areas have been increasing for the last decades in

E. Frascaroli (✉)
Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Viale Fanin, 44, 40127 Bologna, Italy
e-mail: elisabetta.frascaroli@unibo.it

P. Revilla
Misión Biológica de Galicia, Spanish National Research Council (CSIC), Apartado 28, 36080 Pontevedra, Spain

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_17

temperate areas and are used in significant amounts for biogas production in climatic conditions that are far from the optimum for maize. Therefore, maize has an exceptional potential to become a major solution for feed and energy. As a consequence of this wide adaptation, maize biodiversity present in temperate regions is large. However, as maize moves to higher latitudes, it has reduced yield and its distribution is limited by cold springs that hamper biomass production. In temperate areas with short growing seasons, maize yield is also limited by the use of early varieties, unless genotypes tolerant to cold are adopted to allow for the extension of the growing season.

Most high latitude regions of the world are not optimal for maize growth due to cold temperatures and short growth cycles. Indeed, the optimum period for maize in high latitudes is short and the available heat units are limited with biomass and yields not fully expressing their potential. Early planting allows maize to benefit from longer days, escape summer heat (Kucharik 2008), increase growth time and, therefore, reach higher biomass production and yield (Darby and Lauer 2002; Louarn et al. 2008). Moreover, earlier planting also allows maize to mature before autumn rains set (Kucharik 2008). Particularly, in cooler areas, early planting has had significant contributions to yield increases since the beginning of maize breeding. Therefore, early sowing would increase biomass production and stability through cultivation of later varieties and would reduce stress incidence through earlier flowering and grain filling and, consequently, diminishing exposure to drought and parasites. Actually, there is an increasing trend to earlier sowing in the USA (Kucharik 2006). Nevertheless, the main factor limiting early planting of maize is that this practice increases the probability of stress factors affecting seeds and seedlings, particularly in cold conditions, and of opportunistic diseases and pests that can kill the seedlings.

As the limiting factor for early planting and cultivation in high latitudes is cold susceptibility, improving cold tolerance is an important breeding objective. Ideally, cold-tolerant maize can be planted very early in the season and can allow rapid establishment, efficient light interception through fast leaf growth, and finally sustain as long as possible light interception until harvest (Louarn et al. 2010). The major constraints for maize production in temperate areas are cold springs and short growing seasons in higher latitudes, and drought and pests in lower latitudes. Climatic change is modifying the stress patterns, and future perspectives indicate that the climate might become more extreme in large areas; thus, current growing areas are endangered.

In this chapter we review (1) the main issues related to the understanding of the effects of cold stress and the improvement of cold tolerance in maize and (2) the genomic approaches aimed at dissecting the genetic basis of the trait and at selecting genotypes to maximize crop performance in the target environments.

17.2 Understanding and Improvement of Cold Tolerance

17.2.1 Effects of Cold in Maize and Its Evaluation

A large body of literature is available concerning the effects of cold conditions in crops (see Revilla et al. 2005, for a review). Compared to the other main crops, maize can be considered as cold sensitive, and this susceptibility varies along plant growth; actually, critical temperatures for maize development increase as plants grow. Cold stress reduces germination, early crop establishment, and heterotrophic and autotrophic growth. At germination, maize is sensitive to cold temperatures and seedlings are injured by temperatures below 10 °C, which often occur at sowing time in Europe (Takáč 2004). The average temperature recommended for sowing is above 10 °C; when the seed is sown under cooler conditions, the metabolic processes induced by imbibition that are required for activating embryo growth can fail. Actually, cold imbibition alters the membrane of mitochondria and reduces respiration and the subsequent liberation of energy

and mobilization of grain reserves required for heterotrophic growth. In addition, cold stress during imbibition hampers seedling establishment by reducing root growth, which affects water uptake and nutrient flow (Aroca et al. 2001; Hund et al. 2004, 2012). Cold conditions in early stages can affect either the final proportion of germination or days to emergence or both (Revilla et al. 2000; Rodríguezet al. 2008; Huang et al. 2013), and these effects of cold stress on germination depend on kernel size, weight, type and other physical and chemical kernel characteristics. Therefore, varieties susceptible to cold conditions usually show poor stand. The minimum temperature for chilling injury to seeds and seedlings is about 0–5 °C. Cellular and tissue injuries occur at temperatures below 5 °C, quite common in large areas of temperate regions, while growth processes are slowed by temperatures below 15 °C. Temperatures below 10 °C cause cellular and tissue injuries in maize, as well as failures on chlorophyll synthesis and lower photosynthetic rate, including a reduction in the efficiency of photosystem II (PSII, Revilla et al. 2005). When cold stress is more severe, large tissue damage can lead to necrosis (Wise 1995). However, low temperatures generally decrease chlorophyll concentration and activity, and thus productivity and stability (Rodríguez et al. 2007; Strigens et al. 2013) but these detrimental effects are not detected in some genetic backgrounds and ranges of cold temperature (Allam et al. 2016). The main detrimental effects of cold conditions at early stages are thus delayed emergence, reduced chlorophyll content (Figs. 17.1 and 17.2) and photosystem II efficiency (Fig. 17.3) as well as decreased early vigour and biomass synthesis. As secondary effects, damages induced by sub-optimal temperatures imply an increase in oxygen species that in turn cause tissue damage and that may differ in cold-tolerant or susceptible maize genotypes (De Santis et al. 1999). Indeed, plant response to cold stress includes the induction of genes and mechanisms of signal transduction such as abscisic acid, carbon assimilation, photosynthetic efficiency (Kingston-Smith et al. 1997; Leipner et al. 1999; Nguyen et al. 2009; Sobkowiak et al. 2014), lipid composition (Kaniuga et al. 1999) and cell cycle (Rymen et al. 2007).

Fig. 17.1 Cold induced albinism in maize

Fig. 17.2 Determination of chlorophyll content (SPAD units) using a hand-held CCM-200 Chlorophyll Content Meter (OptiSciences, Tyngsboro, Massachusetts, USA)

Physiological analyses of maize plants under cold stress show short-term responses such as decreases in the photosynthetic rate determined as carbon assimilation, transport of photoassimilates, phloem transport, formation of

Fig. 17.3 Determination of fluorescence parameters: QUANTUM efficiency of photosystem II (ΦPSII) by using a portable OS-30p Chlorophyll Fluorometer (Opti-Sciences, Tyngsboro, Massachusetts, USA)

non-photochemical quenching, changes in antioxidant metabolism or concentration of soluble phenols (Leipner and Stamp 2009). Furthermore, cold stress induces changes in cell wall polymers and membrane composition (Boutté and Ggebe 2009; De Santis et al. 2011; Janská et al. 2010, 2011; Moura et al. 2010) that affect stability of the photosynthetic system. Interestingly, some studies also revealed interactions among tolerance to different stresses, showing similar responses between tolerance to cold and to desiccation (Kollipara et al. 2002). These responses are probably mediated by factors such as the *DREB* genes (Cook et al. 2004) which confer tolerance to freezing when overexpressed in maize (Chiappetta et al. 2005). Other examples of cross-talking between stress-reaction mechanisms concern the induction of cold tolerance after imposing mild stress, during acclimation or different treatments (Frascaroli et al. 2005; Frascaroli and Landi 2014; Kingston-Smith et al. 1999; Sobkowiak et al. 2016). Long-term exposure to chilling conditions leads to a reduction in cell division, and in the subsequent tissue growth (Rymen et al. 2007), and to a permanent reduction in photosynthetic activity of leaves that developed under suboptimal temperatures (Kingston-Smith et al. 1999; Nie et al. 1995). This coincides with changes in chloroplast structure and composition affecting thylakoids, photosynthetic complexes and fatty acids (Caffarri et al. 2005; De Santis et al. 1999; Haldimann 1998; Kutik et al. 2004; Nie and Baker 1991; Pinhero et al. 1999; Robertson et al. 1993). Leaves that develop during chilling periods maintain lower radiation efficiency even when temperatures are not stressful. Thus, long-term modifications may be caused by irreversible morphological changes or by signalling originated during stress periods that affect leaves formed after the cold period. As plants develop, even temperatures below 20 °C can be a strong limitation to maize growth and development (Sobkowiak et al. 2016). Low temperature stress can hamper plant functioning in several ways, such as photosynthesis impairment, biochemical changes, damages to tissues and even death. Warrington and Kanemasu (1983) observed that also the leaf number can be affected by low temperature shortly prior to tassel initiation when temperatures lower than 18 °C can reduce the number of leaves. So far, physiological studies have not been able to fully explain the causes of impaired performance of chilling sensitive plants in the short-term and the consequences in the long-term response. As summarized by Marocco et al. (2005), physiological response to chilling stress could be different if it is: (i) mild chilling stress (12–17 °C) in the light, where photosynthesis and growth are reduced and photoprotective mechanisms are induced; (ii) strong chilling stress (2–10 °C) in the light where the rate of transpiration exceeds the rate of water uptake by roots due to inhibition of root hydraulic conductivity and thus cold-induced water stress is induced; and (iii) chilling stress in the dark, mainly associated with changes in gene expression. However, the mechanisms that could be directly responsible for increased cold tolerance are still not defined and we hope for a breakthrough in their understanding from the application of genomic tools.

Furthermore, there are some phenomena involved in response to cold stress that are still scarcely studied, namely the recovery after cold damage (Rodríguez et al. 2013) or the dynamics of stress-induced senescence. Understanding the physiology and the genetic control of cold tolerance and enabling selection for the trait are

impaired by the difficulties in applying cold temperature in controlled as well as in field experiments.

The main limitation for evaluations of cold tolerance consist of the setting up of experiments with controlled, stable and reliable conditions, particularly when large numbers of genotypes must be evaluated. In fact, field trials are unpredictable and heterogeneous due to the changing environmental conditions, while evaluation in controlled conditions may be unrealistic and poorly correlated with field performance. The best compromise to deal with this dilemma is the combination of controlled and field conditions with replications in as many locations and years as possible (Blum 1988; Revilla et al. 2005). Under controlled conditions (Fig. 17.4), there are a variety of approaches for cold tolerance trials, as there are several substrates (soil, peat, sand, perlite or other artificial substrates, and hydroponic solutions) with variable properties; for example, using soil provides similar conditions to those of the field, but soil is heterogeneous and does not allow precise characterizations of roots. Other important factors that need to be considered are ranges of temperatures, light intensities and cycles, moisture and other circumstances, which can become important technical limitations. Normally, cold tolerance is measured as germination at an established low temperature compared to the optimum temperature (Landi et al. 1992) but other determinations are possible and each objective has specific technical requirements. Furthermore, each developmental stage of each crop has different temperature requirements (Revilla et al. 2005).

Fig. 17.4 Evaluation of a maize panel of inbred lines for cold tolerance under controlled conditions

Besides experimental errors due to variable environments and methodologies used for evaluation, stressful conditions can cause large experimental errors (Blum 1988). Another source of error is the origin of the seed; it should be produced in several environments (Revilla et al. 2005) to control the effects of its origin. In order to overcome the above-mentioned drawbacks, precise, high-throughput, and automatic phenotyping using non-destructive techniques in controlled, highly reproducible growth conditions can be adopted. These techniques should provide reliable information about the dynamic of growth and of developmental processes under stress in order to reveal the genetic contribution to the variation for tolerance to cold (Furbank and Tester 2011).

17.2.2 Sources of Cold Tolerance and Breeding

Selection for cold tolerance in modern plant breeding is often based simply on selecting for adaptation to specific environments rather than on addressing cold tolerance directly. However, as previously mentioned, cold stress is often unpredictable because of the vagaries of weather and the relative importance of the trait in achieving the agronomic performance can vary widely. For this reason, variation for yield under stress is largely due to environmental factors, not genetic, and selection based exclusively on yield performance under such conditions would be inadvisable. As a solution, Greaves (1996) suggested to improve performance under cold stress by crossing materials with high yield potential with genotypes adapted to a cold environment. Indeed, maize breeders can benefit from the remarkable genetic diversity available in maize (Gore et al. 2009) and from the considerable ability of maize to respond to selection for adaptation to a wide range of environmental

conditions, as has been demonstrated even for defective mutants as in sweet corn (Ordás et al. 2004, 2006, 2008, 2010). Maize adapted to cold areas is expected to tolerate low temperatures better than maize from warmer origins; particularly for varieties with medium to long growth cycles that cannot be sown late to avoid low spring temperatures. Specifically, maize genotypes from the European Flint race show more cold tolerance than the Corn Belt Dent varieties (Leipner and Stamp 2009; Rodríguez et al. 2010; Strigens et al. 2013) and several previous reports found sources of cold tolerance within European germplasm (Frascaroli and Landi 2013, 2016; Lee et al. 2002; Revilla et al. 2000, 2014; Rodríguez et al. 2010; Verheul et al. 1996). The previous reports indicate that the populations available in temperate germplasm banks are not necessarily more cold tolerant than the elite inbred lines already selected from them or from other sources; consequently, it might not be worthwhile to search favorable alleles for cold tolerance in temperate landraces before we fully exploit the collections of inbred lines. Revilla et al. (2014) evaluated two panels of dent and flint maize inbred lines and found moderate levels of cold tolerance; the dent and flint germplasm most tolerant to cold temperatures were the Northern Flint D171 and the Iodent PH207 groups, respectively. As little genetic variation is present in elite maize germplasm for most agronomic traits, besides cold tolerance, breeders search genetic diversity among exotic maize landraces, which are a rich source of favorable alleles for broadening the genetic base of elite germplasm. Landraces may have great potential, but they also show important handicaps for a fruitful incorporation into breeding programs. Finally, given that the genetic basis is narrow within the resources available in maize collections, there are projects intending to the incorporation of favorable alleles for cold tolerance in collections of teosinte (Hufford et al. 2013). Nevertheless, efficient strategies must be developed to explore this high diversity.

During the last decades, breeders have released a few cold-tolerant genotypes with limited success and marker-assisted selection (MAS) for cold tolerance have not been successfully reported in the literature so far. The complex genetics of cold tolerance and large experimental errors associated to evaluation under stress conditions might explain why traditional breeding programs have reached a ceiling, even though some significant improvements have been obtained by divergent selection for tolerance at germination from the cross between B73 and a Italian inbred line (Frascaroli and Landi 2013, 2017) or for photosynthesis from a Swiss dent maize breeding population (Fracheboud et al. 1999). Furthermore, selection for cold tolerance has often detrimental effects on agronomic traits (Hund et al. 2005; Sezegen and Carena 2009). Development of cold-tolerant varieties would contribute to significant genetic gains in biomass production and yield stability; indeed, both northern cultivation and early sowing of cold-tolerant maize could be a promising strategy for implementing sustainable production of maize in cold areas, even though there is no clear relationship between cold tolerance at early stages of development and yield (Leipner and Stamp 2009). Hopefully, further progress can be expected from the major advances that are ongoing in phenotyping and genotyping techniques and statistical data integration and analysis (Mochida and Shinozaki 2010; Moreno-Risueno et al. 2010).

17.3 Using Genomic Tools for Gene Discovery and for Selection

17.3.1 Genetic Basis of Cold Tolerance

The genetic regulation and the biochemical and physiological basis of cold tolerance in maize are poorly understood even though several genes somehow involved in the expression of the trait have been identified and characterized. Actually, there is a gap between the physiological/biochemical basis of maize response to cold stress and the genetic advances in understanding those underlying the phenotypes observed. From what was described in the previous section, cold

tolerance is a complex polygenic trait with additive, dominance and maternal effects (Revilla et al. 2000; Leipner and Stamp 2009). In addition, Yan et al. (2017) also reported significant heterosis for cold tolerance at seedling stages, in agreement with the significant dominance effects identified by previous authors (Bhosale et al. 2007; Revilla et al. 2000) and with the specific combining ability that has been reported by Hodges et al. (1997). Maternal gene effects detected in maize at the germination and emergence stages (Revilla et al. 2000) appear to restrict their contribution to cold tolerance to the beginning of plant development. Moreover, each growth stage (germination, emergence, early growth, etc.) might be under the control of an independent genetic model because cold tolerance is independently regulated at different growth stages (Hodges et al. 1997; Revilla et al. 2000) and, furthermore, the genetic control of tolerance depends on the material used and the traits studied. No single major genes have been reported to be responsible for cold tolerance in maize; even major genes that affect the agronomic performance at large, such as sweet corn mutants, have not shown sizeable effects on cold tolerance of the genotypes carrying the corresponding mutations (Ordás et al. 2004, 2006, 2008, 2010). In addition to that complexity, significant genotype × environment interaction effects were reported (Fracheboud et al. 2004; Presterl et al. 2007) which makes it even more difficult to determine the genetics of cold tolerance.

With the genomic tools at their disposal, several researchers have made a renewed effort for understanding the genetics of cold tolerance. In particular, since cold tolerance is mainly quantitatively inherited, several studies have concerned quantitative trait loci (QTL) detection. One of the key parameters affected by low temperature is the efficiency of PSII, which is estimated by using fluorescence measurements because chlorophyll-a fluorescence parameters are often used to characterize the stability and function of PSII (Fracheboud et al. 1999). The first results published were mainly focused on the variation in response to cold for chlorophyll content or photosynthesis (Fracheboud et al. 2004; Jompuk et al. 2005; Presterl et al. 2007; Rodríguez et al. 2008, 2014). However, the measurements of fluorescence parameters are not a comprehensive estimate of cold tolerance because many other features are not directly related to the efficiency of PSII; for example, Hund et al. (2004) found that the operating efficiency of PSII was related to seedling dry weight but was not related to root traits. These authors identified 20 QTL for shoot weight and 40 for root weight, length and diameter, and secondary roots, hence concluding that there was a large number of independently inherited loci suitable for the improvement of early seedling growth through better seed vigor and higher rate of photosynthesis. Rodríguez et al. (2014) identified ten QTL associated to maize performance under cold conditions and, through a meta-QTL analysis, these authors identified three genomic regions that regulate the development of maize seedlings under cold conditions and proposed them as useful targets for marker-assisted selection. Yin et al. (2015) reported three major genomic regions associated with chlorophyll fluorescence parameters that were stable and sufficiently strong to be used for marker-assisted selection. Other studies identified QTL associated with cold tolerance at germination (Han et al. 2014; Hu et al. 2016; Li et al. 2018; Wang et al. 2016) or seedling emergence (Liu et al. 2017; Yousef and Juvik 2002). In one case, regions putatively involved in controlling tolerance to cold were detected as a signature of divergent recurrent selection (Frascaroli and Landi 2018).

As revised by Xiao et al. (2017), the exploitation of single nucleotide polymorphisms (SNPs, Gore et al. 2009) and of genotype by sequencing (GBS) have made it possible to search for QTL and quantitative trait polymorphisms by means of genome-wide association studies (GWAS) in maize. Actually, most modern research is focused on GWAS analyses for identifying QTL, with variable success. Several publications have reported promising results on genotyping that could facilitate genomic selection in a short or medium term. GWAS can be

applied to large panels of inbred lines (Revilla et al. 2016) or to more limited panels (Hu et al. 2017); GWAS has also been applied to $F_{2:3}$ generations and testcrosses by Yan et al. (2017).

Revilla et al. (2016) and Strigens et al. (2013) carried out GWAS for cold tolerance in collections of maize inbred lines and identified several QTL explaining low proportions of phenotypic variance for early growth and chlorophyll fluorescence. Strigens et al. (2013) used GWAS to identify QTL for cold tolerance in a panel of maize inbred lines genotyped with 56,110 SNPs and found markers that explained a large proportion of the phenotypic variance for early growth and chlorophyll fluorescence parameters. Revilla et al. (2016) evaluated, under controlled conditions and in the field, two panels for cold tolerance with more than 600 inbred lines and 50 K SNPs and identified the largest number of QTL ever published. They also detected more QTL for cold tolerance in the European flint panel than in the dent panel, most of which were associated with days to emergence and PSII efficiency. Unterseer et al. (2016) performed GWAS for identifying signatures of selection specific to temperate dent and flint pools and identified genes under selective pressure that differed between dent and flint pools. The same genes were also suggested as candidates and investigated for their potential role in the adaptation to specific environments. Reports on panels with fewer inbred lines but more markers referred a reduced number of QTL involved in cold tolerance (Hu et al. 2017).

Several authors have reported a large number of studies on QTL identification and validation based on linkage mapping with biparental populations and GWAS diversity panels (Flint-Garcia et al. 2005). While biparental populations have poor mapping resolution and require validation in subsequent studies, usually by means of near isogenic lines, diversity panels assure high resolution but often are affected by population structure that must be taken into account in order to avoid biases (Vilhjalmsson and Nordborg 2013). More complex mapping populations, like connected populations, have been used to map QTL for tolerance to cold, as in the case of maize at the germination phase (Li et al. 2018) where 650 families allowed to map up to 43 QTL, reduced to three after a meta-analysis of the connected populations was performed. Large panels and recombinant inbred lines (RILs) collections obtained from breeding populations or landraces are a powerful tool that is expected to accelerate high resolution mapping of QTL. Moreover, RILs obtained from multiparental advanced generation intercross (MAGIC), or from landraces, benefit from the high recombination rate of populations that have been multiplied by breeders or farmers for many generations. In both RILs libraries derived from multiparental crosses and GWAS panels, usually multiple alleles segregate making QTL detection more complex. On the other hand, the RILs libraries have the advantage that they are not affected by hidden structures, while the panels usually are, and thus RILs from MAGIC populations have the same benefits as those derived from landraces (Dell'Acqua et al. 2015; Giraud et al. 2014). However, the methods for managing large RILs libraries derived from complex populations require further statistical developments because, in contrast to what happens with biparental populations, the reconstruction of their haplotype mosaic is not straightforward (Huang et al. 2015) and requires a more sophisticated approach in order to reliably define QTL (Ducrocq et al. 2009). Nevertheless, genome-wide prediction has been successful in testcross designs, biparental populations, panels and different types of maize populations (Albrecht et al. 2011; Lehermeier et al. 2014; Riedelsheimer et al. 2013; Rincent et al. 2012).

17.3.2 Gene Discovery

Identification of specific genes for cold tolerance or proxy traits can be pursued by using transcriptomic and proteomic approaches and/or QTL validation and cloning (Salvi et al. 2007; Eichten et al. 2011). In fact, identifying and characterizing cardinal genes for cold tolerance can be accomplished by combining genetic, physiological, structural and functional analyses

with multi-level precision phenotyping and molecular profiling analyses. Sobkowiak et al. (2014) integrated QTL information obtained with biparental mapping populations (Fracheboud et al. 2004; Hund et al. 2005; Jompuk et al. 2005; Leipner and Mayer 2008) with transcriptomic data of cold-treated maize seedlings of cold-tolerant and cold-sensitive parental lines. The results of Sobkowiak et al. (2014) revealed close associations of some differently regulated, cold-responsive expressed regions with major QTL for cold-performance previously identified. This research clearly points out that the cell membrane/cell wall could be potential sensors of low temperature and/or effectors modulating cold response (Fig. 17.5). Earlier information on populations obtained from divergent recurrent selection also detected clear differences in membrane properties correlated with cold tolerance (De Santis et al. 1999, 2011; Tampieri et al. 2011). In subsequent work, Sobkowiak et al. (2016) studied the comparative transcriptomic profile of three maize inbred lines, contrasting in cold-sensitivity both in the field and in a controlled environment. Results revealed that a first response mechanism was related to acclimation of the photosynthetic apparatus, while a second one was related to the cell wall structure, which seems to be modified already at a moderately low temperature thereby facilitating withstanding cold stress. A third response mechanism involved modifications of developmental processes (Sobkowiak et al. 2016). The whole transcriptome analysis performed by means of RNA-sequencing (RNA-seq, Wang et al. 2009) was employed in a study conducted by Li et al. (2016) to identify the potential candidate genes involved in response to freezing stress of plantlets, in tolerant and sensitive inbred lines. Of the 948 genes differentially expressed by tolerant and sensitive lines in response to cold treatment, 30 candidates were validated and thus can be resources for enhanced understanding of early freezing response. Other studies performed on different sets of tolerant/susceptible genotypes or stress combinations contribute to the collection of a large amount of data that altogether will make it possible to obtain a more coherent picture of response to cold (Jończyk et al. 2017; Kollipara et al. 2002; Li et al. 2017; Lu et al. 2017; Shan et al. 2013; Shinozaki et al. 2003; Waters et al. 2017).

The large body of information available on the genome sequence of maize (maizegdb.org) and gene expression (Sekhon et al. 2013) should provide efficient background for further advances on the understanding of the genetics of cold tolerance. Moreover, other available approaches that are expected to clarify this complex situation include epigenetic modifications analysis (Makarevitch et al. 2013). However, to exploit the large amount of data available, bioinformatics tools are needed to allow geneticists and breeders managing the big-data files obtained from the large-scale genomics and breeding experiments (Upadhyay et al. 2017). Several on-line resources can be employed as referred to in Hassani-Pak et al. (2016) that can be integrated in the genome-scale knowledge network. In fact, making good use of the large amount of information available could be the bottleneck for further understanding cold tolerance in the so-called "post-genomic" era. The major challenge nowadays is to store and handle the increasing amount of information present within the genome sequences or even within the transcriptomics data already obtained for an increasing number of plant and crop species. Several tools are available, such as those for sequence analysis and similarity searching, where, by using the genome sequence data as reference set, the candidate genes related to stress are sequenced among various crop and model plants. With this technique it has been possible to evaluate the genetic variation existing for a particular candidate gene and to predict to which stress a specific genomic element is potentially responding (Jaiswal and Usadel 2016). Together with the previous techniques, integrating and visualizing the information present in different databases should permit the access of even novel genes or metabolic pathways reported in various data sets for biotic and abiotic stresses (Hassani-Pak and Rawlings 2017). As an example of the complexity of the genes putatively involved in cold tolerance control, we interrogate

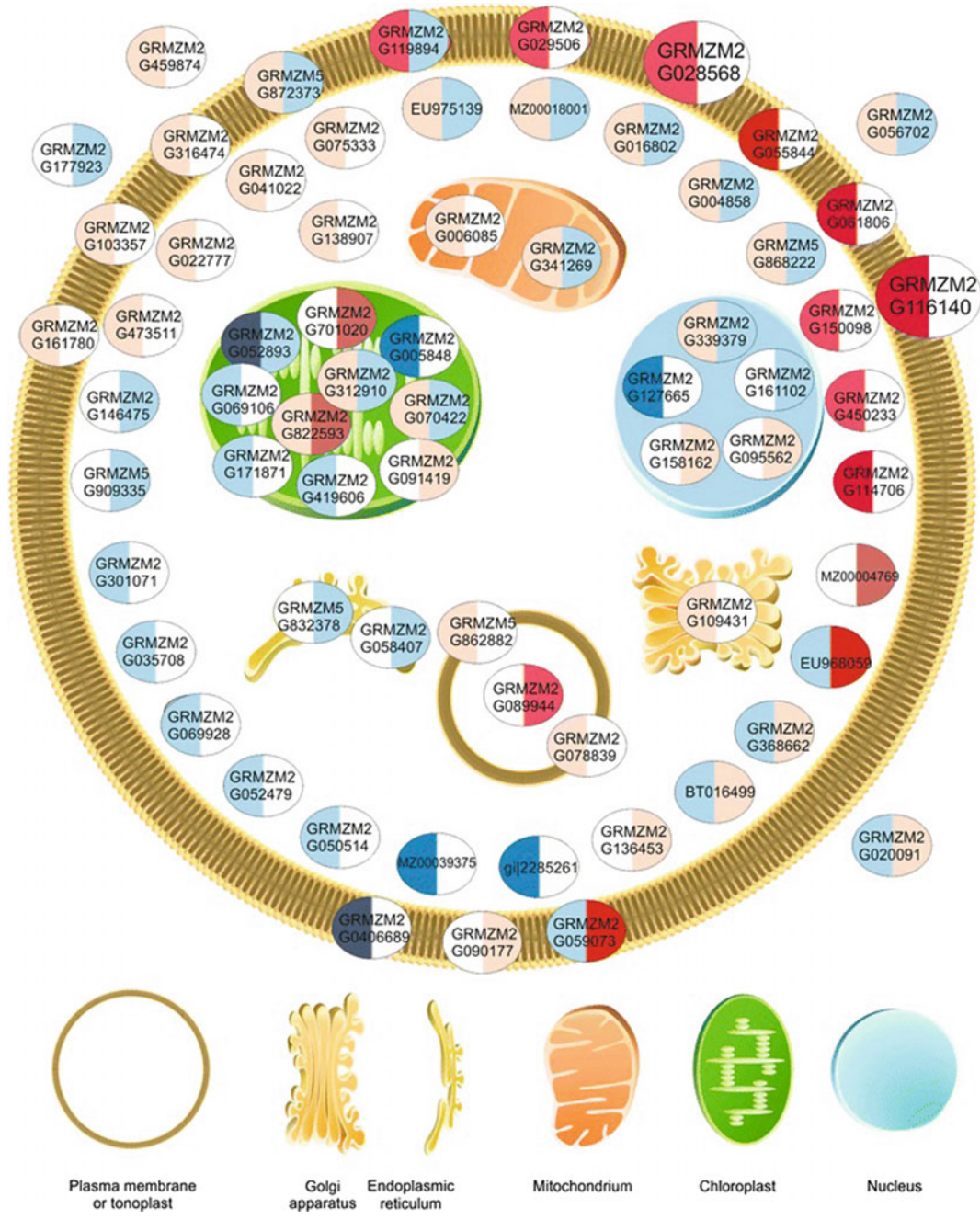

Fig. 17.5 Predicted cellular localization of products of genes showing significant differential expression between lines ETH-DH7 and ETH-DL3 in response to low temperature. In each oval the left-hand half represents the response of the cold-tolerant ETH-DH7 line and the right-hand half the response of the cold-sensitive ETH-DL3 line. Two larger ovals depict proteins related to both cell membrane and cell wall. Colors indicate the cold/control ratio (induction of expression: red; repression: blue; no change: white) and color intensity indicate the magnitude of change (From Sobkowiak et al. 2014)

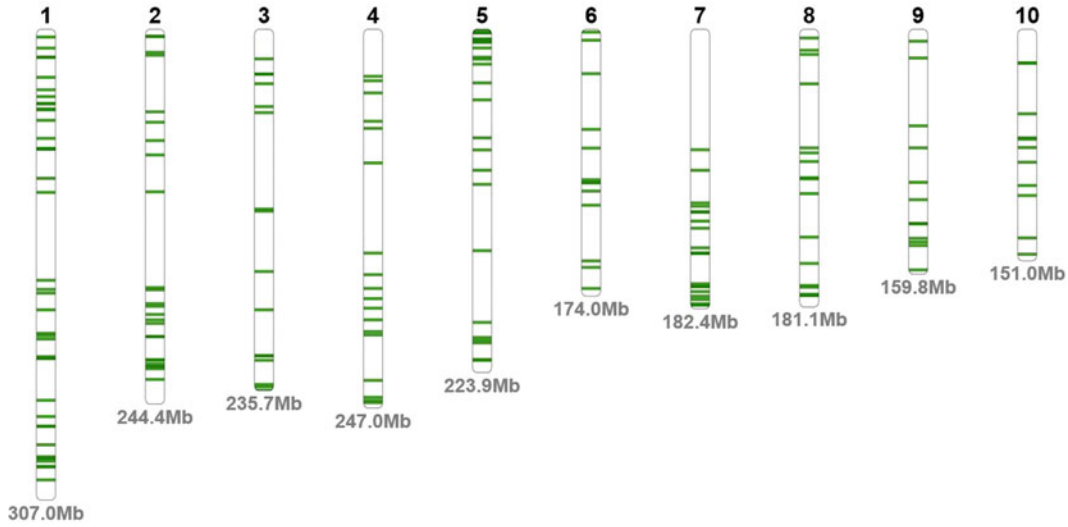

Fig. 17.6 An example of the complexity of the genes putatively involved in cold tolerance control, we interrogate maize databases by means of KnetMiner for the phenotypic trait "cold tolerance" in the whole genome. The chromosomal localization of the first 200 genes only is displayed

maize databases by means KnetMiner (http://knetminer.rothamsted.ac.uk/) for the phenotypic trait "cold tolerance" in the whole genome. The maize database is integrated with information of *Arabidopsis* as a model species to understand the genes function (Hassani-Pak et al. 2016) and in Fig. 17.6 we report the first 200 putative genes recalled under maize with the general trait ontology "cold tolerance" mapping all over the maize map.

Since the publication of the sequence of B73, a large set of genetic and molecular resources have been developed for identification of genes involved in maize performance through forward- and reverse-genetic approaches, including collections of mutants and transformation systems (Nannas and Dawe 2015). The development of RNA-seq and other high-throughput technologies has facilitated unbiased and highly reproducible deep-sequencing of entire transcriptomes (Metzker 2010). Moreover, sequencing transposon insertions with a next generation sequencing method is also available, as McCarty et al. (2013) have released a sequence indexed collection of transposon-induced maize mutants, which comprises thousands of lines with large numbers of insertions. Big efforts are being made to integrate all genomics information of crop and model species into databases to facilitate comparative analysis of genomes, also in response to stress conditions (Naithani et al. 2017; Tello-Ruiz et al. 2018).

17.3.3 Selection with Genomic Tools

Molecular markers and QTL studies have provided a limited benefit on maize breeding for cold tolerance up to now; nevertheless, we now have the best tools ever for improving tolerance to low temperatures, particularly at germination and early development, but also at other sensitive growth stages. Next-generation sequencing technologies allow for deep genotypic characterization of thousands of genotypes that could provide novel approaches for maize breeding (Langridge and Fleury 2010; Meyer et al. 2007; Riedelsheimer et al. 2012).

GWAS enables the identification of the QTL, candidate gene and even causative SNPs underlying the phenotype of interest. These approaches require the use of highly diverse panels of genotypes, as Nested Association Mapping (NAM) panels, obtaining RILs and introgression

library lines in reference inbreds derived from selected inbred donors, and similar resources that should be densely genotyped and extensively phenotyped. The combination of large phenotypic and genomic databases on one hand will facilitate the dissection of the QTLome (Salvi and Tuberosa 2015) and the cloning of the genes for improving cold tolerance, and on the other hand, with genomic selection (GS) proposed by Meuwissen et al. (2001), will accelerate the selection of genotypes carrying favorable alleles at loci undetectable by means of mapping approaches.

Current research is expected to yield basic insights on cold tolerance and applied aspects such as identification and characterization of useful genes for improving maize tolerance to cold stress with advanced information on the affected molecular and physiological processes. This information could be useful to obtain allele sequences of functionally characterized genes from which functional motifs affecting plant phenotype can be identified and may be used as functional markers (Andersen and Lübberstedt 2003; Brenner et al. 2013).

With the enhancement of knowledge of plant genomics, and especially of sequence technology, plant science will shift its focus from "explanatory" to "predictive". Indeed, so far most research in plant genomics has been addressed towards understanding the molecular basis of biological processes or phenotypic traits. Conversely, from a plant breeding perspective, in the case of traits quantitatively inherited, the main interest is in predicting the breeding value of the best genotypes by using molecular information for more time- and cost-efficient breeding schemes. The possibility to predict the optimal genotype or genotype combination based on genomic information would greatly enhance the efficiency of plant breeding programs. This is especially true in the case of GS, where dense marker coverage is instrumental to maximize the number of QTL whose effects will be captured by markers (Abdel-Ghani and Lübberstedt 2013). Simulation and empirical studies revealed that GS efficiency is superior to that of MAS (Bernardo and Yu 2007), especially for traits that are difficult to measure such as cold tolerance, particularly in the field. Most advanced prediction techniques (Montesinos-Lòpez et al. 2018) are promising in the perspective of implementation of selection for cold tolerance in combination with other agronomic traits.

17.4 Perspectives and Challenges

Recent advances in genomics, metabolomics and bioinformatics allow us to confront the challenge of improving maize for complex traits such as cold tolerance in order to increase production and stability under stress conditions that are getting worse with climatic change. Basically, the strategies for selection of cold tolerance can involve:

- Intensive genotyping of large association panels of inbred lines or RILs populations obtained from biparental progenies, MAGIC populations, NAM designs or local populations;
- High-precision, next-generation phenotyping in controlled growth chambers and platforms, and multi-environment field trials combining agronomic, morphological, physiological, and biochemical data;
- Statistical models for (i) the identification of genomic regions involved in cold tolerance and exploitation of favorable alleles, in the case of MAS or (ii) the prediction of the breeding value of non-phenotyped genotypes, as in the case of GS.

References

Abdel-Ghani AH, Lübberstedt T (2013) Parent selection—usefulness and prediction of hybrid performance. In: Lübberstedt T, Varshney RK (eds) Diagnostics in plant breeding. Springer, Dordrecht, pp 349–368

Albrecht T, Wimmer V, Auinger H-J, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön CC (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Allam M, Revilla P, Djemel A, Tracy WF, Ordás B (2016) Identification of QTLs involved in cold tolerance in sweet x field corn. Euphytica 208:353–365

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Aroca R, Tognoni F, Irigoyen JJ, Sanchez-Diaz M, Pardossi A (2001) Different root low temperature response of two maize genotypes differing in chilling sensitivity. Plant Physiol Bioch 39:1067–1073

Bernardo R, Yu J (2007) Prospects for genome-wide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Bhosale SU, Rymen B, Beemster GTS, Melchinger AE, Reif JC (2007) Chilling tolerance of central European maize lines and their factorial crosses. Ann Bot 100:1315–1321

Blum A (1988) Plant Breeding for stress environments. CRC Press, Inc., Boca Ratón

Boutté Y, Grebe M (2009) Cellular processes relying on sterol function in plants. Curr Opin Plant Biol 12:705–713

Brenner EA, Beavis WD, Andersen JR, Lübberstedt T (2013) Prospects and limitations for development and application of functional markers in plants. In: Lübberstedt T, Varshney RK (eds) Diagnostics in plant breeding. Springer, Dordrecht, pp 329–346

Caffarri S, Frigerio S, Olivieri E, Righetti PG, Bassi R (2005) Differential accumulation of Lhcb gene products in thylakoid membranes of *Zea mays* plants grown under contrasting light and temperature conditions. Proteomics 5:758–768

Chiappetta L, Tomes D, Xu D, Sivasankar S, Sanguineti MC, Tuberosa R (2005) *DREB1* overexpression improves tolerance to low temperature in maize. In: Tuberosa R, Phillips RL, Gale M (eds) Proceedings of the international congress "In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution" Bologna, Italy, Avenue media, Bologna, Italy, pp 533–544

Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci USA 101:15243–15248

Darby HM, Lauer JG (2002) Planting date and hybrid influence on corn forage yield and quality. Agron J 94:281–289

De Santis A, Landi P, Genchi G (1999) Changes of mitochondrial properties in maize seedlings associated with selection for germination at low temperature. Fatty acid composition, cytochrome c oxidase, and adenine nucleotide translocase activities. Plant Physiol 119:743–754

De Santis A, Frascaroli E, Baraldi E, Carnevali F, Landi P (2011) The activity of the plant mitochondrial inner membrane anion channel (PIMAC) of maize populations divergently selected for cold tolerance level is differentially dependent on the growth temperature of seedlings. Plant Cell Physiol 52:193–204

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J, Frascaroli E, Churchill GA, Inzé D, Morgante M, Pè ME (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome Biol 16:16

Ducrocq S, Giauffret C, Madur D, Combes V, Dumas F, Jouanne S, Coubriche D, Jamin P, Moreau L, Charcosset A (2009) Fine mapping and haplotype structure analysis of a major flowering time quantitative trait locus on maize chromosome 10. Genetics 183:1555–1563

Eichten SR, Foerster J, de Leon N, Kai Y, Yeh C-T, Liu S, Jeddeloh J, Schnable PS, Kaeppleer SM, Springer NM (2011) B73-Mo17 near isogenic lines (NILs) demonstrate dispersed structural variation in maize. Plant Physiol 156:1679–1690

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064

Fracheboud Y, Jompuk C, Ribaut JM, Stamp P, Leipner J (2004) Genetic analysis of cold-tolerance of photosynthesis in maize. Plant Mol Biol 56:241–253

Fracheboud Y, Haldimann P, Leipner J, Stamp P (1999) Chrlorophyll fluorescence as a selection tool for cold tolerance of photosynthesis in maize (*Zea mays* L.). J Exp Bot 338:1533–1540

Frascaroli E, Casarini E, Conti S (2005) Response of maize inbred lines to a defoliation treatment inducing tolerance to cold at germination. Euphytica 145:295–303

Frascaroli E, Landi P (2013) Divergent selection in a maize population for germination at low temperature in controlled environment: study of the direct response, of the trait inheritance and of correlated responses in the field. Theor Appl Genet 126:733–746

Frascaroli E, Landi P (2014) Inheritance of the responses to a defoliation treatment affecting cold tolerance in maize. Euphytica 197:319–329

Frascaroli E, Landi P (2016) Cold tolerance in field conditions, its inheritance, agronomic performance and genetic structure of maize lines divergently selected for germination at low temperature. Euphytica 209:771–788

Frascaroli E, Landi P (2017) Registration of Maize Inbred Line Bo23 with high cold tolerance and agronomic performance for early sowing. J Plant Regist 11:172–177

Frascaroli E (2018) Landi P (2018) Signatures of divergent selection for cold tolerance in maize. Euphytica 214:80. https://doi.org/10.1007/s10681-018-2163-x

Furbank RT, Tester M (2011) Phenomics—technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16:635–644

Giraud H, Lehermeier C, Bauer E, Falque M, Segura V, Bauland C, Camisan C, Campo L, Meyer N, Ranc N, Schipprack W, Flament P, Melchinger AE, Menz M, Moreno-González J, Ouzunova M, Charcosset A, Schön CC, Moreau L (2014) Linkage disequilibrium with linkage analysis of multiline crosses reveals

different multiallelic QTL for hybrid performance in the flint and dent heterotic groups of maize. Genetics 198:1717–1734

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Greaves JA (1996) Improving suboptimal temperature tolerance in maize—the search for variation. J Exp Bot 47:307–323

Haldimann P (1998) Low growth temperature-induced changes to pigment composition and photosynthesis in *Zea mays* genotypes differing in chilling sensitivity. Plant Cell Environ 21:200–208

Han Z, Ku L, Zhang Z, Zhang J, Guo S, Liu H, Zhao R, Ren Z, Zhang L, Su H, Dong L, Chen Y (2014) QTLs for seed vigor-related traits identified in maize seeds germinated under artificial aging conditions. PLoS ONE 9(3):e92535

Hassani-Pak K, Castellote M, Esch M, Hindle M, Lysenko A, Taubert J, Rawlings C (2016) Developing integrated crop knowledge networks to advance candidate gene discovery. Appl Transl Genom 11:18–26

Hassani-Pak K, Rawlings C (2017) Knowledge discovery in biological databases for revealing candidate genes linked to complex phenotypes. J Integr Bioinform 14 (1). Retrieved 11 Apr 2018 from https://doi.org/10.1515/jib-2016-0002

Hodges DM, Andrews CJ, Johnson DA, Hamilton RI (1997) Sensitivity of maize hybrids to chilling and their combining abilities at two developmental stages. Crop Sci 37:850–856

Hu G, Li Z, Lu Y, Li C, Gong S, Yan S, Li G, Wang M, Ren H, Guan H, Zhang Z, Qin D, Chai M, Yu J, Li Y, Yang D, Wang T, Zhang Z (2017) Genome-wide association study identified multiple genetic loci on chilling resistance during germination in maize. Sci Rep 7 N 10840

Hu S, Lübberstedt T, Zhao G, Lee M (2016) QTL mapping of low-temperature germination ability in the maize IBM Syn4 RIL population. PLoS ONE 11(3): e0152795

Huang J, Zhang J, Li W, Hu W, Duan L, Feng Y, Qiu F, Yue B (2013) Genome-wide association analysis of ten chilling tolerance indices at the germination and seedling stages in maize. J of Integr Plant Biol 55:735–744

Huang BE, Verbyla KL, Verbyla AP, Raghavan C, Singh VK, Gaur P, Leung H, Varshney RK, Cavanagh CR (2015) MAGIC populations in crops: current status and future prospects. Theor Appl Genet 128:999–1017

Hufford MB, Lubinksy P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J (2013) The genomic signature of crop-wild introgression in maize. PLoS Genet 9:e1003477

Hund A, Fracheboud Y, Soldati A, Frascaroli E, Salvi S, Stamp P (2004) QTL controlling root and shoot traits of maize seedlings under cold stress. Theor Appl Genet 109:618–629

Hund A, Frascaroli E, Leipner J, Jompuk C, Stamp P, Fracheboud Y (2005) Cold tolerance of the photosynthetic apparatus: pleiotropic relationship between photosynthetic performance and specific leaf area of maize seedlings. Mol Breed 16:321–331

Hund A, Reimer R, Stamp P, Walter A (2012) Can we improve heterosis for root growth of maize by selecting parental inbred lines with different temperature behaviour? Phil Trans R Soc B 367:1580–1588

Jaiswal P, Usadel B (2016) Plant pathway databases. Plant Bioinform Methods Protoc 2016:71–87

Janská A, Aprile A, Zámečník J, Cattivelli L, Ovesná J (2011) Transcriptional responses of winter barley to cold indicate nucleosome remodelling as a specific feature of crown tissues. Funct Integr Genomics 11:307–325

Janská A, Maršík P, Zelenková S, Ovesná J (2010) Cold stress and acclimation-what is important for metabolic adjustment? Plant Biol 12:395–405

Jompuk C, Fracheboud Y, Stamp P, Leipner J (2005) Mapping of quantitative trait loci associated with chilling tolerance in maize (*Zea mays* L.) seedlings grown under field conditions. J Expl Bot 56:1153–1163

Jończyk M, Sobkowiak A, Trzcinska-Danielewicz J, Skoneczny M, Solecka D, Fronk J, Sowiński P (2017) Global analysis of gene expression in maize leaves treated with low temperature. II. Combined effect of severe cold (8 °C) and circadian rhythm. Plant Mol Biol 95:279

Kaniuga Z, Saczynska V, Miskiewcz E, Garstka M (1999) The fatty acid composition of phosphatidylglycerol and sulfoquinovosyldiacylglycerol of *Zea mays* genotypes differing in chilling susceptibility. J Plant Physiol 154:256–263

Kingston-Smith AH, Harbinson J, Foyer CH (1999) Acclimation of photosynthesis, H_2O_2 content, and antioxidants in Maize (*Zea mays*) grown at suboptimal temperatures. Plant Cell Environ 22:1071–1083

Kingston-Smith AH, Harbinson J, Williams J, Foyer CH (1997) Effect of chilling on carbon assimilation, enzyme activation, and photosynthetic electron transport in the absence of photoinhibition in maize leaves. Plant Physiol 114:1039–1104

Kollipara KP, Saab IN, Wych RD, Lauer MJ, Singletary GW (2002) Expression profiling of reciprocal maize hybrids divergent for cold germination and desiccation tolerance. Plant Physiol 129:974–992

Kucharik CJ (2006) A multidecadal trend of earlier corn planting in the central USA. Agron J 98:1544–1550

Kucharik CJ (2008) Contribution of planting date trends to increased maize yields in the central United States. Agron J 100:328–336

Kutik J, Holá D, Koová M, Rothova O, Haisel D, Wilhelmová N, Tichá I (2004) Ultrastructure and dimensions of chloroplasts in leaves of three maize (*Zea mays* L.) inbred lines and their F_1 hybrids grown under moderate chilling stress. Photosynthetica 42:447–455

Landi P, Frascaroli E, Lovato A (1992) Divergent full-sib recurrent selection for germination at low temperature in a maize population. Euphytica 64:21–29

Langridge P, Fleury D (2010) Making the most of 'omics' for crop improvement. Trends Biotechnol 29:33–40

Lee EA, Staebler MA, Tollenaar M (2002) Genetic variation in physiological discriminators for cold tolerant-early autotrophic phase of maize development. Crop Sci 42:1919–1929

Lehermeier C, Kramer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger AE, Menz M, Meyer N Moreau L, Moreno-González J, Ouzunova M, Pausch H, Ranc N, Schipprack W, Schönleben M, Walter H, Charcosset A, Schön CC (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. Genetics 198:3–16

Leipner J, Fracheboud Y, Stamp P (1999) Effect of growing season on the photosynthetic apparatus and leaf antioxidative defenses in two maize genotypes of different chilling tolerance. Environ Exp Bot 42:129–139

Leipner J, Mayer E (2008) QTL mapping in maize seedlings reveals little relevance of C4 cycle enzymes and antioxidants for genotypic differences in chilling tolerance of photosynthesis. Maydica 53:269–277

Leipner J, Stamp P (2009) Chilling stress in maize seedlings. In: Bennetzen JL, Hake SC (eds) Handbook of maize: its biology. Springer, Heidelberg

Li P, Cao W, Fang H, Xu S, Yin S, Zhang Y, Lin D, Wang J, Chen Y, Xu C, Yang Z (2017) Transcriptomic profiling of the maize (*Zea mays* L.) leaf response to abiotic stresses at the seedling stage. Front Plant Sci 8:290

Li X, Wang G, Fu J, Li L, Jia G, Ren L, Lubberstedt T, Wang G, Wang J, Gu R (2018) QTL mapping in three connected populations reveals a set of consensus genomic regions for low temperature germination ability in *Zea mays* L. Front Plant Sci 9:65

Li Z, Hu G, Liu X, Zhou Y, Li Y, Zhang X, Yuan X, Zhang Q, Yang D, Wang T, Zhang Z (2016) Transcriptome sequencing identified genes and gene ontologies associated with early freezing tolerance in maize. Front Plant Sci 7:1477

Liu H, Zhang L, Wang J, Li C, Zeng X, Xie S, Zhang Y, Liu S, Hu S, Wang J, Lee M, Lübberstedt T, Zhao G (2017) Quantitative trait locus analysis for deep-sowing germination ability in the maize IBM Syn10 DH population. Front Plant Sci 8:813

Louarn G, Andrieu B, Giauffret C (2010) A size-mediated effect can compensate for transient chilling stress affecting maize (*Zea mays*) leaf extension. New Phytol 187:106–118

Louarn G, Chenu K, Fournier C, Andrieu B, Giauffret C (2008) Relative contributions of light interception and radiation use efficiency to the reduction of maize productivity under cold temperatures. Funct Plant Biol 35:885–899

Lu X, Zhou X, Cao Y, Zhou M, McNeil D, Liang S, Yang C (2017) RNA-seq analysis of cold and drought desponsive transcriptomes of *Zea mays* ssp. *mexicana* L. Front. Plant Sci 8:136

Makarevitch I, Eichten SR, Briskine R, Waters AJ, Danilevskaya ON, Meeley RB, Myers CL, Vaughn MW, Springer NM (2013) Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. Plant Cell 3:780–793

Marocco A, Lorenzoni C, Fracheboud Y (2005) Chilling stress in maize. Maydica 50:571–580

McCarty DR, Latshaw S, Wu S, Suzuki M, Hunter CT, Avigne WT, Koch KE (2013) Mu-seq: sequence-based mapping and identification of transposon induced mutations. PLoS ONE 8(10):e77172

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Törjék O, Fiehn O, Eckhardt Ä, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate revealed by canonical correlation analysis in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 104:4759–4764

Mochida K, Shinozaki K (2010) Genomics and bioinformatics resources for crop improvement. Plant Cell Physiol 51:497–523

Montesinos-López OA, Montesinos-López A, Crossa J, Montesinos-López JC, Mota-Sanchez D, Estrada-González F, Gillberg J, Singh R, Mondal S, Juliana P (2018) Prediction of multiple-trait and multiple-environment genomic data using recommender systems. G3 (Bethesda) 8:131–147

Moreno-Risoeno MA, Busch W, Benfey P (2010) Omics meets networks —using systems approaches to infer regulatory networks in plants. Curr Opin Plant Biol 13:126–131

Moura JC, Bonine CA, de Oliveira Fernandes Viana J, Dornelas MC, Mazzafera P (2010) Abiotic and biotic stresses and changes in the lignin content and composition in plants. J Integr Plant Biol 52:360–376

Naithani S, Preece J, D'Eustachio P, Gupta P, Amarasinghe V, Dharmawardhana PD, Wu G, Fabregat A, Elser JL, Weiser J, Keays M, Fuentes AM, Petryszak R, Stein LD, Ware D, Jaiswal P (2017) Plant Reactome: a resource for plant pathways and comparative analysis. Nucleic Acids Res. 4; 45(D1): D1029–D1039

Nannas NJ, Dawe RK (2015) Genetic and genomic toolbox of Zea mays. Genetics 199:655–669

Nguyen HT, Leipner J, Stamp P, Guerra-Peraza O (2009) Low temperature stress in maize (*Zea mays* L.) induces genes involved in photosynthesis and signal transduction as studied by suppression subtractive hybridization. Plant Physiol Biochem 47:116–122

Nie G-Y, Baker NR (1991) Modifications to thylakoid composition during development of maize leaves at low growth temperature. Plant Physiol 95:184–191

Nie G-Y, Robertson EJ, Freyer MJ, Leech RM, Baker NR (1995) Response of the photosynthetic apparatus in

maize leaves grown at low temperature on transfer to normal growth temperature. Plant Cell Environ 18:1–12

Ordás B, Malvar RA, Soengas P, Ordás A, Revilla P (2004) *Sugary1* inbreds to improve sugary enhander1 hybrids of sweet corn for adaptation to cold areas with short growing seasons. Maydica 49:279–288

Ordás B, Padilla G, Malvar RA, Ordás A, Rodríguez VM, Revilla P (2006) Cold tolerance improvement of sugary enhancer1 hybrids of sweet corn. Maydica 51:567–574

Ordás B, Revilla P, Ordás A, Malvar RA (2008) Hybrids sugary × sugary enhancer of sweet corn: a valuable option for cool environments. Sci Hortic 118:111–114

Ordás B, Rodríguez VM, Romay MC, Malvar RA, Ordás A, Revilla P (2010) Adaptation of super-sweet corn to cold conditions: mutant by genotype interaction. J Agric Sci 148:401–405

Pinhero RG, Palianth G, Yada Y, Murr DP (1999) Chloroplast membrane organization in chilling-tolerant and chilling-sensitive maize seedlings. J Plant Physiol 155:691–698

Presterl T, Ouzunova M, Schmidt W, Möller EM, Röber FK, Knaak C, Ernst K, Westhoff P, Geiger HH (2007) Quantitative trait loci for early plant vigour of maize grown in chilly environments. Theor Appl Genet 114:1059–1070

Revilla P, Butrón A, Cartea ME, Malvar RA, Ordás A (2005) Breeding for cold tolerance. In: Ashraf M, Harris PJC (eds) Abiotic stresses. Plant resistance through breeding and molecular approaches. The Haworth Press, Inc., USA pp. 301–398

Revilla P, Malvar RA, Cartea ME, Butrón A, Ordás A (2000) Inheritance of cold tolerance at emergence and during early season growth in maize. Crop Sci 40:1579–1585

Revilla P, Rodríguez VM, Ordás A, Rincent R, Charcosset A, Giauffret C, Melchinger AE, Schön CC, Bauer E, Altmann T, Brunel D, Moreno-González J, Campo L, Ouzunova M, Laborde J, Álvarez Á, Ruíz de Galarreta JI, Malvar RA (2014) Cold tolerance in two large maize inbred panels adapted to European climates. Crop Sci 54:1981–1991

Revilla P, Rodriguez VM, Ordas A, Rincent R, Charcosset A, Giauffre C, Melchinger AE, Schon CC, Bauer E, Altmann T, Brunel D, Moreno-Gonzalez J, Campo L, Ouzunova M, Alvarez A, Ruiz de Galarreta JI, Laborde J, Malvar RA (2016) Association mapping for cold tolerance in two large maize inbred panels. BMC Plant Biol 16:127

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220

Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194:493–503

Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, Rodriguez VM, Moreno-Gonzales J, Melchinger AE, Bauer E, Schön C-C, Meyer N, Giauffret C, Bauland C, Jamin P, Laborde J, Monod H, Flament P, Charcosset A, Moreau L (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). Genetics 192:715–728

Robertson EJ, Baker NR, Leech RM (1993) Chlorophyll thylakoid protein changes induced by low growth temperature in maize revealed by immunocytology. Plant Cell Environ 16:809–818

Rodríguez VM, Butrón A, Malvar RA, Ordás A, Revilla P (2008) Quantitative trait loci for cold tolerance in the maize IBM population. Int J Plant Sci 169:551–556

Rodríguez VM, Butrón A, Rady MOA, Soengas P, Revilla P (2014) Identification of QTLs involved in the response to cold stress in maize (*Zea mays* L.). Mol Breed 33:363–371

Rodríguez VM, Malvar RA, Butrón A, Ordás A, Revilla P (2007) Maize populations as sources of favorable alleles to improve cold tolerant hybrids. Crop Sci 47:1779–1786

Rodríguez VM, Romay MC, Ordás A, Revilla P (2010) Evaluation of the European Maize (*Zea mays* L.) germplasm under cold conditions. Gen Res Crop Evol 57:329–335

Rodríguez VM, Velasco P, Garrido JL, Revilla P, Ordás A, Butrón A (2013) Genetic regulation of cold-induced albinism in the maize inbred line A661. J Expl Bot 64:3657–3667

Rymen R, Fiorani F, Kartal F, Vandepoele K, Inzé D, Beemster GTS (2007) Cold nights impair leaf growth and cell cycle progression in maize through transcriptional changes of cell cycle genes. Plant Physiol 143:1429–1438

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski JA, Tingey SV, Miao GH, Phillips RL, Tuberosa R (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Natl Acad Sci USA 104:11376–11381

Salvi S, Tuberosa R (2015) The crop QTLome comes of age. Curr Opin Biotechnol 32:179–185

Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, Buell CR, de Leon N, Kaeppler SM (2013) Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE 8:e61005

Sezegen B, Carena MJ (2009) Divergent recurrent selection for cold tolerance in two improved maize populations. Euphytica 167:237–244

Shan X, Li Y, Jiang Y, Jiang Z, Hao W, Yuan Y (2013) Transcriptome profile analysis of maize seedlings in response to high-salinity, drought and cold stresses by deep sequencing. Plant Mol Biol Rep 31:1485

Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought

and cold stress responses. Curr Opin Plant Biol 6:410–417

Sobkowiak A, Jończyk M, Adamczyk J, Szczepanik J, Solecka D, Kuciara I, Hetmańczyk K, Trzcinska-Danielewicz J, Grzybowski M, Skoneczny M, Fronk J, Sowiński P (2016) Molecular foundations of chilling-tolerance of modern maize. BMC Genom 17:125

Sobkowiak A, Jonczyk M, Jarochowska E, Biecek P, Trzcinska-Danielewicz J, Leipner J, Fronk J, Sowinski P (2014) Genome-wide transcriptomic analysis of response to low temperature reveals candidate genes determining divergent cold-sensitivity of maize inbred lines. Plant Mol Biol 85:317–331

Strigens A, Freitag NM, Gilbert X, Grieder C, Riedelsheimer C, Schrag TA, Messmer R, Melchinger AE (2013) Association mapping for chilling tolerance in elite flint and dent maize inbred lines evaluated in growth chamber and field experiments. Plant Cell Environ 36:1871–1887

Takáč T (2004) The relationship of antioxidant enzymes and some physiological parameters in maize during chilling. Plant Soil Environ 50:27–32

Tampieri E, Baraldi E, Carnevali F, Frascaroli E, De Santis A (2011) The activity of plant inner membrane anion channel (PIMAC) can be performed by a chloride channel (CLC) protein in mitochondria from seedlings of maize populations divergently selected for cold tolerance. J Bioenerg Biomembr 43:611–621

Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A, Wei S, Preece J, Geniza MJ, Jiao Y, Lee YK, Wang B, Mulvaney J, Chougule K, Elser J, Al-Bader N, Kumari S, Thomason J, Kumar V, Bolser DM, Naamati G, Tapanari E, Fonseca N, Huerta L, Iqbal H, Keays M, Munoz-Pomer Fuentes A, Tang A, Fabregat A, D'Eustachio P, Weiser J, Stein LD, Petryszak R, Papatheodorou I, Kersey PJ, Lockhart P, Taylor C, Jaiswal P, Ware D (2018) Gramene 2018: unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res 46:D1181–D1189

Unterseer S, Pophaly SD, Peis R, Westermeier P, Mayer M, Seidel MA, Haberer G, Mayer KFX, Ordas B, Pausch H, Tellier A, Bauer E, Schoen CC (2016) A comprehensive study of the genomic differentiation between temperate Dent and Flint maize. Genome Biol 17 N 137

Upadhyay J, Joshi R, Singh B, Bohra A, Vijayan R, Bhatt M, Bisht S, Wani SH (2017) Application of bioinformatics in understanding of plant stress tolerance. In: Hakeem K, Malik A, Vardar-Sukan F, Ozturk M (eds) Plant bioinformatics. Springer, Cham, pp 347–374

Verheul MJ, Picatto C, Stamp P (1996) Growth and development of maize (*Zea mays* L.) seedlings under chilling conditions in the field. Eur J Agron 5:31–43

Vilhjalmsson BJ, Nordborg M (2013) The nature of confounding in genome-wide association studies. Nat Rev Genet 14:1–2

Wang B, Zhang Z, Fu Z, Liu Z, Hu Y, Tang J (2016) Comparative QTL analysis of maize seed artificial aging between an immortalized F_2 population and its corresponding RILs. Crop J 4:30–39

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Warrington IJ, Kanemasu ET (1983) Corn growth response to temperature and photoperiod. I. seedling emergence tassel initiation and anthesis. Agron J 75:749–754

Waters A, Makarevitch I, Noshay J, Burghardt L, Hirsch CN, Hirsch CD, Springer N (2017) Natural variation for gene expression responses to abiotic stress in maize. Plant J 89:706–717

Wise RR (1995) Chilling-enhanced photooxidation. The production, action and study of reactive oxygen species produced during chilling in the light. Photosynth Res 45:79–97

Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017) Genome-wide association studies in maize: praise and stargaze. Mol Plant 10:359–374

Yan J, Wu Y, Li W, Qin X, Wang Y, Yue B (2017) Genetic mapping with testcrossing associations and $F_{2:3}$ populations reveals the importance of heterosis in chilling tolerance at maize seedling stage. Scientific Reports 7 N 3232

Yin Z, Qin Q, Wu F, Zhang J, Chen T, Sun Q, Zhang Y, Wang H, Deng D (2015) Quantitative trait locus mapping of chlorophyll a fluorescence parameters using a recombinant inbred line population in maize. Euphytica 205:25–35

Yousef GG, Juvik JA (2002) Enhancement of seedling emergence in sweet corn by marker-assisted backcrossing of beneficial QTL. Crop Sci 42:96–104

High-Oil Maize Genomics

18

Xiaohong Yang and Jiansheng Li

Abstract

High-oil maize is a crop developed by artificial selection. Maize oil is high in energy and levels of polyunsaturated fatty acids, which makes high-oil maize a popular resource for food, feed, and bioenergy. Multiple high-oil germplasm resources have been developed, mainly including the Illinois High-Oil (IHO), Alexho synthetic, and Beijing High-Oil (BHO) populations. Yet, the molecular mechanisms underlying oil biosynthesis and accumulation are not well understood. Historically, quantitative genetic approaches like QTL mapping, and recently developed association mapping, have been utilized to understand the genetic architecture of oil biosynthesis and accumulation in maize kernels. Subsequently, the genes related to oil biosynthesis and accumulation were cloned by homolog-based cloning, position cloning, and association mapping. These cloned genes are involved in the oil metabolic pathway, transcription factors, and regulators controlling oil storage organ. Favorable alleles of most cloned genes for kernel oil-related traits were mined and are promising targets for improving oil quantity and quality in maize. The successful and effective transformation of the favorable allele of *DGAT1-2* into elite maize hybrids confirms the effectiveness of these favorable alleles in the manipulation of oil quantity and quality.

18.1 Introduction

High-oil maize is a crop developed by artificial selection. In general, the kernel oil content in commercial maize hybrids is around 4.5% on a dry weight basis, while it is over 6% in high-oil maize hybrids. Maize oil mainly accumulates in the embryo, which contains about 85% of the total lipids, much more than lipids in endosperm (3%) and aleurone (12%) (Shen and Roesler 2017). Chemically, maize oil is comprised of approximately 11% palmitic acid (C16:0), 2% stearic acid (C18:0), 24% oleic acid (C18:1), 62% linoleic acid (C18:2), and 1% linolenic acid (C18:3) (Lambert 2001).

Maize oil is highly valued for both animal feed and human food. The caloric content of oil is 2.25 times greater than that of starch on a weight basis, and livestock feeding studies have shown that the growth rate, feed efficiency, and productivity of livestock improves with increasing oil content within the maize kernel (Han et al. 1987; Benitez et al. 1999; Lambert et al. 2004). In addition, the small amount of linolenic acid

X. Yang · J. Li (✉)
State Key Laboratory of Plant Physiology and Biochemistry, National Maize Improvement Center of China, China Agricultural University, Beijing 100193, China
e-mail: lijiansheng@cau.edu.cn

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_18

and high levels of natural antioxidants in maize oil, such as phytosterols and vitamin E, enhance oil stability that makes it suitable for high-temperature frying applications (Weber 1987; Lambert 2001). Furthermore, the relatively high level of total unsaturated fatty acids (~87%), especially linoleic acid, makes maize oil relatively healthy, as unsaturated fatty acids have been shown to lower blood cholesterol levels (Ramsden et al. 2013). Taken together, maize oil is high in polyunsaturated fatty acids and low in linolenic acid, making it a desirable vegetable oil. Therefore, improving the quantity and quality of maize kernel oil content is an important target for maize breeding.

18.2 Development of High-Oil Maize Germplasm

The first long-term artificial selection for oil content was initiated with the open-pollinated variety Burr's White at the University of Illinois in 1896 (Hopkins 1899; Dudley and Lambert 2004; Lucas et al. 2013). It has become a "textbook" example of the power of artificial selection. The original objective was to determine whether the chemical composition of the maize kernel could be changed by selection. To begin the experiment, Hopkins analyzed 163 ears of the open pollinated variety Burr's White for oil content. The 24 highest ears and 12 lowest ears in oil content were selected to form Illinois High-Oil (IHO) and Illinois Low-Oil (ILO) strains, respectively. Recurrent selection in these two populations has been performed every year since 1896 except three years during World War, and considerable progress in response to cycles of selection has been observed in the strains (Fig. 18.1a). After 110 generations of selection, the oil content in the IHO strain increased from 4.7 to 24.4%, while the oil content in the ILO strain reached to a low limit (<1.0%), which was too low to accurately measure.

Following 48 generations of forward selection, reverse selection was initiated in each of the two strains to create the Reverse High-Oil (RHO) and the Reverse Low-Oil (RLO) strains, respectively (Leng 1962; Dudley and Lambert 2004). After 7 generation of selections in RHO strain, selection was again reversed to initiate the Switchback High-Oil (SHO) strain (Dudley and Lambert 2004). Response to selection continued to be observed in all three populations (Fig. 18.1a). After 110 generations of selection for the IHO and the ILO strains, the oil content decreased from 13.5 to 3.6% in the RHO, increased from 0.8 to 6.0% in the RLO, and reached 22.2% in the SHO.

Gains from selection for decreased oil content in the ILO strain are no longer observed, suggesting that the ILO strain has reached a lower biological limit for oil content owing to its poor

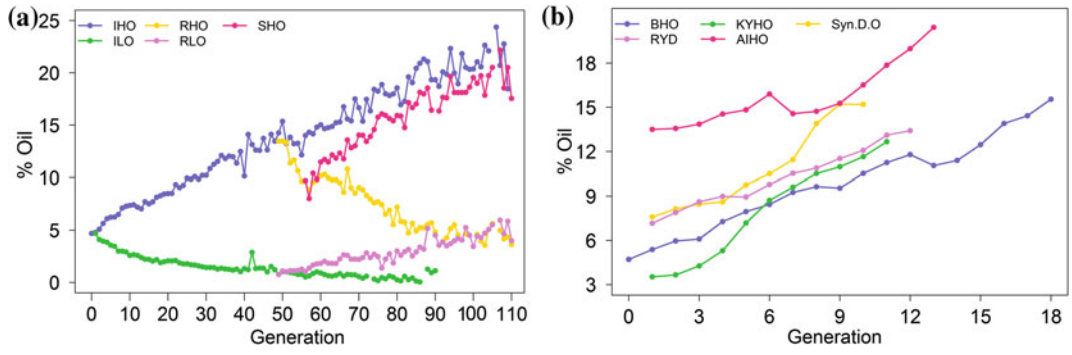

Fig. 18.1 Averaged oil content in maize kernel at different selection generations for high-oil populations. **a** IHO, ILO, RHO, RLO, and SHO populations. **b** BHO, KYHO, Syn.D.O, RYD and AIHO populations. Data from the studies by Dudley and Lambert (2004), Song and Chen (2004), Wang et al. (2009), Lucas et al. (2013)

germination rate and lack of genetic diversity in recent generations (Dudley and Lambert 2004; Moose et al. 2004). Conversely, an upper limit has not been reached for oil content in the IHO and SHO strains despite their oil content reaching up over 20%. This indicates that abundant genetic diversity still existed in the IHO and SHO strains and continued selection for higher oil content is possible in maize.

To broaden the germplasm base of high-oil maize, Alexander initiated maize breeding programs after 1956 (Miller et al. 1981; Mišević and Alexander 1989; Lambert et al. 2004). He developed several synthetics, including Alexho, AE (Alexho Elite), UHO (Ultra High-Oil), DO (Disease Oil Synthesic), ARYD (Arnel'S Reid Yellow Dent), RSSSCHO (Stiff-Stalk Synthetic), and BS10HO (Iowa 2-Ear). After four to 28 generations of selection, the oil content increase ranged from 4.0–4.8 to 7.0–22.1% (Table 18.1). Among these populations, Alexho, AE, and UHO were derived from Alexho synthetic, which was developed by mixing 43 open-pollinated maize varieties, two crosses involving exotic germplasm by corn belt inbreds, and four Illinois chemical strains, containing high-oil, low-oil, high-protein, and low-protein strains. Details of the selection procedures and statistical evaluations have been reported previously by Lambert et al. (2004). In addition, five high-oil populations with different genetic backgrounds, BHO (Beijing High-Oil), AIHO (Alexho-IHO High-Oil), Syn.D.O. (Synthesic Disease Oil), RYD (Reid Yellow Dent), and KYHO (KY High-Oil), were developed by Tongming Song at China Agricultural University (Song and Chen 2004). In the early 1980s, Tongming Song visited Illinois University and returned with seeds of $ASKC_{23}$, $IHOC_{80}$, DO, and ARYO (Lambert et al. 2004). Among these five high-oil populations, AIHO, Syn.D.O., and RYD trace their origin to Illinois synthetics, while BHO and KYHO were developed from Chinese synthetic, Zhongzong No.2, and 14 Chinese elite inbreds, respectively. The selection procedure used to develop these five populations was basically the same for all populations and has been described in detailed by Song and Chen (2004). In brief, 100–120 ears were selected in each cycle and then 100 kernels of each ear were analyzed by NMR. The highest three to four kernels from each ear were saved and mixed. Finally, 300–340 kernels of each population were planted, divided into A and B plots, and pollinated with bulked pollen for each other. After seven to 18 cycles of selection, the oil content increased by 6.3–10.0% across the populations (Fig. 18.1b). Taking the BHO population as an example, the kernel oil content increased from 4.7 to 15.6% after 18 selection cycles.

18.3 QTL Mapping for Oil Content and Composition

Maize oil content is a complex quantitative trait controlled by multiple loci or genes, with the broad-sense heritability ranging from 62.3 to 98.0% (Table 18.2). Precisely locating and characterizing these functional loci will facilitate

Table 18.1 Selection for oil content in Alexho, DO, ARYD, RSSSCHO, BS10HO populations

High-oil population[a]	Cycles of selection[b]	Oil content (%)[b]	
		C_0	Cn
Alexho	28	4.58	22.06
DO	8	4.20	9.50
ARYD	7	4.04	9.12
RSSSCHO	5	4.70	7.10
BS10HO	4	4.84	6.99

[a]*DO* Disease Oil Synthesic; *ARYD* Arnel'S Reid Yellow Dent; *RSSSCHO* Stiff-Stalk Synthetic; *BS10HO* Iowa 2-Ear
[b]Data from studies by Miller et al. (1981), Lambert et al. (2004)

Table 18.2 Summary of QTL for oil content

Populations[a]	Population size	Marker[b]	Mapping method[c]	QTL No.	R^2 (%)[d]	H^2 (%)[e]	References
IHP(76) × ILP(76) F_2S_1	100	RFLP	SMA	25	43.0	74.0	Goldman et al. (1994)
IHP(70) × IHO(70) F_1RM_{10}:S_2	500	SNP	SMR	50	50.0	96.0	Laurie et al. (2004)
By804 × B73 F_2, $F_{2:3}$	450	SSR	CIM	11, 15		56.7, 62.3	Song et al. (2004)
L-20-01 × L-02-03 F_2	408	SSR	CIM	13	26.6	98.0	Mangolin et al. (2004)
IHP(70) × IHO(70) F_1RM_{10}:S_2 PS	500	SNP	SIM	51		93.0	Clark et al. (2006)
IHP(70) × IHO(70) F_1RM_{10}:S_2 TC	500		SIM	54		94.0	
IHP(70) × IHO(70) $F_1RM_7S_2$ PS	500	SNP	SIM	70		89.0	Dudley et al. (2007)
IHP(70) × IHO(70) $F_1RM_7S_2$ TC	500	SNP	SIM	63		77.0	
IHO(90) × B73 BC_1S_1	150	RFLP and SSR	CIM	5	46.9	86.0	Wassom et al. (2008a, b)
IHO(90) × B73 BC_1S_1 TC	150	RFLP and SSR	CIM	3	17.5	91.0	
By804 × B73 $F_{2:3}$	298	SSR	CIM	6	53.0	81.8	Zhang et al. (2008)
By804 × B73 RIL	245	SSR	MCIM	9	55.3	92.5	Yang et al. (2010)
L-14-4B/L-08-05F $F_{2:3}$	250	SSR	MCIM	16	30.9	89.0	Môro et al. (2012)
8984 × GY220 RIL	282	SSR	CIM	12		76.0	Yang et al. (2012)
8622 × GY220 RIL	263	SSR	CIM	14		63.0	
P53/178 RIL	498	SSR	ICIM	13	14.0–28.3	80.9	Zhang et al. (2015)
Yu82 × Yu87-1 RIL	208	SNP	MCIM	4		78.0	Wang et al. (2016)
Yu82 × Shen137 RIL	197	SNP	MCIM	4		82.0	
Yu87-1 × Zong3 RIL	223	SNP	MCIM	4		77.0	
Teosinte NILs	961	SNP	JSR	6	45.0	94.0	Karn et al. (2017)

[a]*RM* random mating; *PS* Per Se; *TC* test cross; *RIL* recombinant inbred line; *NIL* near isogenic lines
[b]*RFLP* restriction fragment length polymorphism; *SSR* simple sequence repeat; *SNP* single-nucleotide polymorphism
[c]*SMA* single marker analysis; *SMR* stepwise multiple regression with MAXR/BIC; *CIM* composite interval mapping; *SIM* simple interval mapping; *mCIM* joint analysis in multiple environments based on composite interval mapping; *ICIM* the inclusive composite interval mapping; *MCIM* a mixed linear model based on composite interval mapping; *JSR* joint step regression
[d]The total phenotypic variation explained by all QTL for oil content identified in a population
[e]The broad-sense heritability

the genetic improvement of oil quantity and quality via marker-assisted selection (MAS) or biotechnology aided breeding.

QTL mapping is a classical way to dissect the genetic architecture of complex quantitative traits. The first QTL mapping for oil content was performed by using an F_2 population derived from crosses of IHP × ILP cycle 76 (Goldman et al. 1994). One hundred F_2 S_1 derived families were genotyped by using 100 polymorphic RFLP loci spaced throughout the maize genome, and the oil content in these families was evaluated in replicated trials. A total of 25 loci in 13 genomic regions were identified to be significantly associated with oil content by single factor analysis of variance and multiple regression analyses, with a high proportion of loci showing additive effects. Since this first QTL study, multiple studies have been conducted to identify QTL associated with oil content in maize kernels by using different QTL mapping methods in various populations (Table 18.2). The populations were developed from crosses between high-oil and regular maize (e.g., By804/B73 segregating and recombinant inbred line (RIL) populations), regular and regular maize (e.g., Yu87-1/Yu82 RIL), tropical and temperate maize (e.g., L-14-4B/L-08-05F segregating population), and maize and teosinte (e.g., Teosinte near isogenic line (NIL)). These populations with different genetic backgrounds facilitated deeper understanding of the genetic architecture of oil content in maize.

The number of QTL associated with oil content in each population ranged from 3 to 70 (Table 18.2). In a large randomly mated population developed from a cross between IHO cycle 70 × ILO cycle 70, approximately 50 QTL for oil content were identified, which accounted for over 50% of phenotypic variation (Laurie et al. 2004; Clark et al. 2006; Dudley et al. 2007). These results agreed well with earlier predictions of many minor genetic factors controlling oil content (Dudley 1977). This suggested that oil content is controlled by a large number of genes with small but additive effects. In contrast, using segregating or RIL populations, a relatively small number of QTL were detected, accounting for less than 56% of the total phenotypic variation in oil content (Goldman et al. 1994; Mangolin et al. 2004; Song et al. 2004; Wassom et al. 2008a; Zhang et al. 2008, 2015; Yang et al. 2010, 2012; Môro et al. 2012; Wang et al. 2016). Some of the identified QTL explaining over 10% of the phenotypic variation, indicating that in some genetic backgrounds variation for oil content may be controlled by a few QTL with very large effects.

Epistasis, the interaction between alleles from two to more genetic loci (Fisher 1918), may also play an important role in the quantitative variation of oil content in maize. It can partly explain why the percent of phenotypic variation explained by all identified QTL in most studies (<56%) is lower than the broad-sense heritability (>62.3%) (Table 18.2). The method for detecting epistatic QTL is immature. Still, Dudley (2008) and Yang et al. (2010) both found the importance of epistatic interaction in contributing to oil content variation. The analysis of two-way epistatic interactions in 500 S_2 lines from the crosses of IHO cycle 70 × ILO cycle 70 and of IHP cycle 70 × ILP cycle 70 demonstrated that the number of markers associated only with significant epistatic effects ranged from 56.8 to 64.1% of the total number of markers significant for either an interaction effect or from single marker analysis (SMA) (Dudley 2008). Yet, only two pairs of epistatic interactions were detected for oil content in the B73 × By804 RIL population, with each pair of epistatic interaction explaining 1.3 and 3.8% of the phenotypic variation. The mapped epistatic QTL comprised two types of interactions: interactions between two QTL with additive effects, and interactions between a QTL with additive effect and a locus without significant additive effect. Taken together, the phenotypic variation of oil content is likely due mainly to additive effects, which is consistent with the dissection of genetic variance components for oil content (Moreno-Gonzales et al. 1975; Miller et al. 1981).

Maize oil is composed largely of triacylglycerol, and the quantity of triacylglycerol is determined by the amount of five fatty acid compositions: palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2), and linolenic (18:3) acids (Lambert 2001), which is mainly reflected

by oil biosynthesis. Compared with QTL mapping for oil content, fewer studies have been performed for fatty acid composition, and those that did look at composition focused on the five key fatty acid compositions due to their high quantity. For each fatty acid composition in each population, 2–17 loci were identified, and in total explained 15.4–67.0% of the phenotypic variation (Alrefai et al. 1995; Wassom et al. 2008b; Yang et al. 2010). Among the identified QTL, some QTL accounted for extremely high percent of the phenotypic variation, such as *pal9* for palmitic acid (42%) and *lin6-1* for linoleic acid (48.3%) in the B73 × By804 RIL population (Yang et al. 2010). Notably, a considerable portion of QTL for fatty acid composition are co-localized with QTL for oil content, suggesting that some loci increase oil content by increasing fatty acid composition.

The capacity of the triacylglycerol storage organ is the second key factors affecting oil accumulation in kernels. Triacylglycerol is stored within maize kernel, which consists of endosperm and embryo. The endosperm, which is composed mainly of starch component, accounts for 80% of kernel mass, whereas the embryo accounts on average for only 10% of kernel mass (Val et al. 2009). However, about 85% of total kernel oil is located in the embryo, compared to only 3% in the endosperm (Shen and Roesler 2017). Thus, the ratio of embryo-to-endosperm weight can partly determine the accumulation of oil, because oil content is negatively correlated with starch content. The IHO experiment demonstrated that embryo size was also associated with oil content in maize kernel (Dudley and Lambert 2004). Therefore, the identification of QTL for traits involved in physical characteristics of the kernel, including the ratio of embryo-to-endosperm weight and embryo size, will allow causal variants of oil content to be identified.

Up to now, only one study was conducted in the B73 × By804 RIL population to detect QTL for traits related to the capacity of the triacylglycerol storage organ, involving in embryo-to-endosperm weight ratio, embryo volume, embryo width, embryo length, and embryo width-to-length ratio (Yang et al. 2012). Phenotypic correlations showed oil content was positively correlated with all traits mentioned above (r = 0.21–0.81), primarily indicating the contribution of embryo related traits to oil content. For each trait, 3–10 QTL were identified, with a total explained phenotypic variation ranging from 15.5 to 52.9%. Out of 12 QTL for oil content, 8 QTL were co-localized with QTL for embryo related traits, validating their molecular contribution to kernel oil content. For example, *qKO1-1* on chromosome 1 had the largest effect on kernel oil content in the B73 × By804 RIL population, and the traits related to embryo size were the key factors for increasing oil content at this locus. Similarly, for *qKO9* on chromosome 9, it is very likely that the embryo-to-endosperm weight ratio contributes the increase of oil content associated with this locus.

Taken together, the QTL mapping for oil content and compositions indicates that the number of genetic factors controlling oil content is likely around 50, the genetic variance components for oil content were largely determined by additive effects. Finally, some genetic factors for embryo-related traits and fatty acid compositions contribute to the increase of oil content.

18.4 Association Mapping for Oil Content and Composition

Although QTL mapping is a powerful and popular approach for identifying the genes or loci affecting natural phenotypic variation, the resolution provided by QTL mapping is low (10–30 cm) unless huge mapping populations are used (Salvi and Tuberosa 2005). Fine mapping of QTL to a more precise genetic position is generally required to clone the underlying gene, which is a resource- and time-consuming process (Xiao et al. 2017). In addition, the large and complex maize genome, more than 60% of which consists of transposable elements (Jiao et al. 2017), further slows the progress in QTL fine-mapping. Association mapping using diverse populations provides another strategy to effectively fine map QTL that takes advantage of

historical recombination events that lead to the rapid decay of linkage disequilibrium (LD) (Flint-Garcia et al. 2003). Maize is an ideal crop for association mapping due to its great genetic diversity and rapid LD decay (Yan et al. 2011), which definitely increase the mapping resolution.

The first genome-wide association studies (GWAS) for oil-related traits were performed by DuPont Pioneer company. A total of 8,590 SNPs were tested for association with oleic acid composition in 553 maize inbreds by Kolmogorov–Smirnov test (Beló et al. 2008). A fatty acid desaturase, *FAD2*, was identified to be responsible for natural variation in oleic acid content. Further resequencing of *FAD2* identified a SNP (G/T) in exon 2, affecting residue 230 (Ser/Ala) and resulting in a polarity change in the amino acid, which was significantly associated with oleic acid composition (Li et al. 2013).

With the explosive development of next-generation sequencing (NGS) technologies and the release of the maize B73 reference genome coupled with GWAS, there are new opportunities to understand the genetic architecture of oil content and composition. Two platforms were used to genotype an association panel consisting of 508 maize inbred lines (Li et al. 2012, 2013; Fu et al. 2013). One is the commercial maizeSNP50 beadchip, which contained 56,110 SNPs (Li et al. 2012, 2013). The other is RNA sequencing (RNA-seq), which produced 1.03 million SNPs and expression data of over 28,000 genes for the developing kernels at 15 DAP in a subset of 368 lines (Fu et al. 2013). The advantage of the RNA-seq genotyping platform is to obtain abundant and informative SNPs from expressed regions of the genome and to simultaneously monitor the expression level of each of the analyzed loci in the context of its biological function.

Using these genotypic data and phenotypic data in four environments, a total of 74 loci were identified to be significantly associated with oil content and composition. Based on the annotation of the 74 most likely genes in these detected loci, about 1/3 were implicated in lipid metabolism, 1/3 were classified as transcription factors, stress response, and enzymes involved in other biological pathways, and the remaining 1/3 have unknown function. Among the 74 loci, 26 loci were significantly associated with oil content, explaining up to 83% of the phenotypic variation. In contrast, only a few loci were identified for each oil component trait, with the locus number of each trait ranging from 1 to 7 (Fig. 18.2). This finding is consistent with the nature biosynthesis of oil content and composition, as oil content is the final product mixture of a complex biosynthesis pathway, while oil composition represents the intermediate products of a pathway. Subsequently, the favorable alleles of the 58 loci associated with oil content and composition (excluding those loci associated only with derived ratios of compositions) were counted in high-oil and regular maize inbreds, respectively. As expected, high-oil maize has more favorable alleles than regular maize. It indicates favorable allele accumulation is one major route for increasing oil content during the selection of high-oil lines.

In addition to the association between marker and oil content and composition, expression QTL (eQTL) for 67 out of 74 loci with available expression data were also performed as the differences in expression level may account for a significant proportion of differences in the traits,

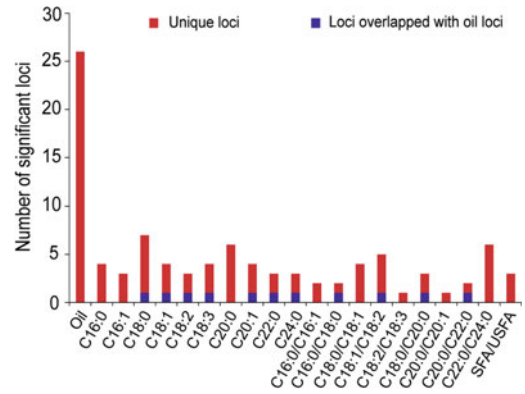

Fig. 18.2 Summary of loci for oil content and compositions identified by GWAS in a maize association panel containing 508 inbred lines. C16:0, palmitic acid; C16:1, palmitoleic acid; C18:0, stearic acid; C18:1, oleic acid; C18:2, linoleic acid; C18:3, linolenic acid; C20:0, arachidic acid; C20:1, gadoleic acid; C22:0, behenic acid; C24:0, lignoceric acid; SFA, saturated fatty acid; USFA, unsaturated fatty acid

especially for quantitative traits. Among these loci, 41 loci defined clear eQTL, 14 of the 41 genes also correlated directly with the phenotypic variation of the target trait, and 18 genes were correlated with a related trait, all at the $P < 0.01$. This strongly suggests that at least some of the genes affect phenotypic variation via transcriptional regulation.

According to the two studies of GWAS on oil content and composition, it is true that GWAS is a high-power and high-resolution QTL analysis. However, it is very difficult to correct for false positive association signals that are caused by population structure (Flint-Garcia et al. 2003) and to identify rare variants that are associated with the traits of interest (Xiao et al. 2017). To solve these issues, a nested association mapping (NAM) population was developed, which consists of a set of 25 recombinant inbred line (RIL) populations derived from crosses between the maize inbred line "B73" and other 25 genetically diverse inbred lines (Yu et al. 2008; McMullen et al. 2009). This population exhibits a powerful potential to thoroughly dissect the genetic architecture of complex quantitative traits due to the joint use of historical and recent recombination events and purified population structure. Cook et al. (2012) used the NAM population to dissect the genetic architecture of oil content. Joint stepwise regression identified 22 oil QTL in the NAM population, which collectively explained 70% of the total variation. Using a subsampling-based multiple SNP model, 135 SNPs with resample model inclusion probability (RMIP) ≥ 0.05 were significantly associated with oil content. Of these loci, only 5 loci were co-localized with the loci significantly associated with oil content identified in an association mapping panel by Li et al. (2013), indicating the remaining loci detect in NAM population might be rare allele in the association mapping panel. These results enriched our knowledge about the genetic architecture of oil content in maize kernel.

18.5 Genes and Favorable Alleles for Oil Content and Composition

Though QTL mapping and association mapping of oil content and composition in maize, studies have reported multiple QTL, yet the molecular basis of oil QTL remains largely unknown. Up to now, only 9 genes for oil-related traits have been cloned by homology-based cloning, positional cloning, or association mapping (Table 18.3). These genes involved functions in the oil biosynthesis, transcription factor, and regulator in embryo size.

Table 18.3 List of genes cloned in previous studies

Genes	Cloning methods	Descriptions	References
ZmSAD1	Homolog-based cloning	Stearoyl-ACP desaturase	Merlo et al. (1998)
ZmLEC1	Homolog-based cloning	LEAFY COTYLEDON1, a HAP3 subunit of the CCAAT-binding factor	Shen et al. (2010)
ZmWRI1	Homolog-based cloning	WRINKLED1, a transcription factor containing two AP2 domains	Shen et al. (2010)
DGAT1-2	Positional cloning	acyl-CoA:diacylglycerol acyltransferase	Zheng et al. (2008)
ZmFatB	Positional cloning and association mapping	A type B fatty acyl-ACP thioesterase	Li et al. (2011), Zheng et al. (2014)
ZmGE2	Homolog-based cloning and positional cloning	GIANT EMBRYO, a cytochrome p450 protein	Zhang et al. (2012)
FAD2	Association mapping	Fatty acid desaturase	Beló et al. (2008)
ACP	Association mapping	acyl carrier protein	Li et al. (2013)
LACS	Association mapping	Long-chain acyl-CoA synthetase	Li et al. (2013)

An extensive genetic knowledge of the storage oil biosynthetic pathways, including both the genes and their regulatory factors, has been generated in the model plant *Arabidopsis* (Beisson et al. 2003; Baud and Lepiniec 2010). Thus, it is available to clone *Arabidopsis* orthologs in maize by homology-based cloning. *ZmSAD1*, encoding stearoyl-ACP desaturase, is the first cloned gene for fatty acid compositions based on the conserved amino acids in the castor (Merlo et al. 1998). It converts stearic to oleic acid through the insertion of a double bond between C9 and C10. As expected, overexpression of *ZmSAD1* in maize kernels reduced stearic acid content by 1.57%, and consequently reduced the saturated to unsaturated fatty acid ratio by 20.40% (Du et al. 2016). Furthermore, *ZmSAD1* was resequenced in a maize association panel to mine favorable alleles (Han et al. 2017). One nonsynonymous single-nucleotide polymorphism in exon 3 and one 5-bp insertion/deletion in the 3′ untranslated region (UTR) was shown to contribute to the natural variation in C18:0/C18:1.

Following *ZmSAD1*, *ZmLEC1,* and *ZmWRI1* have been identified as two key transcription factors involved in the regulation of oil accumulation based on the conserved amino acids in *Arabidopsis* (Shen et al. 2010). *ZmLEC1* encodes a HAP3 subunit of the CCAAT-binding factor, while *ZmWRI1* encodes a transcription factor containing two AP2 domains. Expression of the *WRI1* gene in *Arabidopsis* has been shown to be under the direct control of the transcription factor *LEC1* (Baud et al. 2007). Overexpression of *ZmLEC1* increased kernel oil content by as much as 48% but reduced seed germination and leaf growth in maize. For *ZmWRI1*, overexpression resulted in an oil increase similar to overexpression of *ZmLEC1* without affecting germination, seedling growth, or grain yield. These results highlight *ZmWRI1* as a promising target for increasing oil production in crops. There are two *ZmWRI1* genes in maize, *ZmWRI1a* on chromosome 2 and *ZmWRI1b* chromosome 4. Further, candidate-gene association mapping found that a 2,000-bp InDel in the 3'UTR of *ZmWRI1a* was significantly associated with oil content (Li et al. 2013).

In addition to the genes identified by homology-based cloning, three more genes were cloned by positional cloning. *DGAT1-2*, underlying *qHO6*, a major QTL for oil content on chromosome 6, is the first gene cloned by positional cloning for oil-related traits in maize (Zheng et al. 2008). It encodes an acyl-CoA diacylglycerol:acyltransferase and catalyzes the final step of oil synthesis by adding a third acyl chain to diacylglycerol and yielding triglyceride. The comparison of *DGAT1-2* cDNA in *qHO6* NILs revealed that a phenylalanine insertion (F469) in the last exon of *DGAT1-2* was responsible for increasing both oil content and oleic acid concentration (Zheng et al. 2008). The effect of this insertion was subsequently validated by resequencing and candidate-gene association analysis (Chai et al. 2011). In addition, overexpression of the high-oil *DGAT1-2* allele increases oil and oleic acid contents by up to 41 and 107%, respectively (Zheng et al. 2008).

Subsequently, *ZmGE2* for embryo size and *ZmfatB* for palmitic acid composition were cloned by positional cloning combined with association mapping or homology-based cloning (Li et al. 2011; Zhang et al. 2012). In the By804 × B73 RIL population, *qHO1* and *qPal9* is the largest QTL for oil content and palmitic acid composition, respectively (Yang et al. 2010). *ZmGE2*, falling within *qHO1*, encodes a cytochrome P450 protein, homologous to rice *GIANT EMBRYO* (*GE*) regulating rice embryo size (Koh et al. 1996; Cahoon et al. 2003; Zhang et al. 2012). Followed by fine mapping and association mapping, a 247-bp transposable element (TE) insertion in the 3′ UTR of *ZmGE2* was identified to be associated with increase in the embryo-to-endosperm ratio and kernel oil content. This finding suggests that oil content in maize kernels can be modified by some genes controlling embryo size, agreeing with the co-localization of QTL for oil content and embryo-related traits. For *qPal9*, the region was firstly mapped to a 90-kb region by fine mapping and association mapping, in which there is only one candidate gene, *ZmfatB*, encoding acyl-acyl carrier protein (ACP) thioesterase (Li et al. 2011). Candidate-gene association mapping

further revealed that an 11-bp insertion in the last exon of *ZmfatB* decreases palmitic acid production, consequently leading to an optimization of the ratio of saturated to unsaturated fatty acids while having no effect on total oil content. Similarly, a recent report found that a single-nucleotide (G) insertion in the 6th exon of *ZmfatB*, which creates a premature stop codon, reduced levels of palmitic acid (Zheng et al. 2014).

Furthermore, three genes encoding the key enzymes or proteins in oil biosynthesis were identified by GWAS and the functional sites were mined by candidate-gene association mapping, including an 8-bp InDel in the 3' UTR of ACP, a nonsynomynous SNP in the exon of *FAD2*, and a 146/472-bp InDel in 3' UTR of *LACS* (Table 18.3). *ACP* encodes an acyl carrier protein, and functions as the mobile carrier of the growing fatty acid chain in each cycle reaction of fatty acid synthesis (Shintani and Ohlrogge 1994). *FAD2* encodes fatty acid desaturase, catalyzing oleic acid to produce linoleic acid (Mikkilineni and Rocheford 2003). *LACS* encodes a long-chain Acyl-CoA synthetase and activates fatty acyl chains to fatty acid CoAs and participates in the last step of fatty acid synthesis and in cutin, polyester and wax synthesis (Li-Beisson et al. 2010). All the functional sites for these three genes, together with six genes identified by fine mapping and homolog-based cloning, are highlighted to be promising targets for improved oil quantity and quality in maize.

18.6 Application of Favorable Alleles in the Improvement of Oil Quantity and Quality

In maize kernels, the oil content is generally negatively correlated with grain yield, which might have two main reasons. One is the negative correlation between kernel oil and starch content (Song and Chen 2004; Clark et al. 2006), and the other is the accumulation of unfavorable alleles for grain yield and other agronomic traits during the artificial selection of high-oil maize (Dudley 1977; Mišević and Alexander 1989). Therefore, an effective method to increase oil content without altering grain yield and other agronomic traits is a key step in breeding high-oil maize varieties. With the increase in cloned genes and mapped quantitative trait loci (QTL) and the availability of inexpensive and reliable marker systems, marker-assisted selection provides an alternative to improve target traits such as oil content of maize kernels.

As mentioned above, the favorable allele (F469) of *DGAT1-2* improves oil content in maize kernel, but influences neither the proportion of the seed occupied by the embryo nor yield production (Zheng et al. 2008; Chai et al. 2011), suggesting a promising application for improvement of oil content in maize kernel via marker-assisted selection. Chai et al. (2011) developed a functional marker based on the 3-bp InDel in the last exon of *DGAT1-2*. Then, Hao et al. (2014) transferred the favorable allele of *DGAT1-2* from the high-oil inbred line, By804, into the parents of Zhengdan958, Zheng58, and Chang7-2, using marker-assisted backcrossing. Two improved inbred lines, Zheng58-qHO6 and Chang7-2-qHO6, were developed through six generations of backcrosses guided by molecular markers. An approximately 260-kb fragment from the donor parent was transferred into recurrent lines, and over 99% of the recurrent genomes were recovered. Both of the improved inbred lines showed increased absolute oil content of roughly 1.0% without a change in grain weight. Consequently, the oil content in improved Zhengdan958-qHO6, crossed from Chang7-2-qHO6 to Zheng58-qHO6, reached 4.5%, with increases in absolute and relative content of 0.7 and 18.0%, respectively, compared with the original Zhengdan958. The grain yield of the improved Zhengdan958-qHO6 ranged from 5,928 to 11,826 kg/ha in ten environments, similar to the original Zhengdan958. This study provides a practical example of the feasibility of improving quantitative traits by transferring desirable alleles using marker-assisted backcrossing.

18.7 Prospective

The complexity of oil biosynthesis and accumulation in maize kernels demands multiple approaches for understanding the underlying mechanisms. Fortunately, technologies have enabled global investigations of the associations between genomic variations and oil content and composition, and the expression level of oil-related genes. An enhanced understanding of the regulation of oil biosynthesis and accumulation and their relationships to oil quantity and quality are expected to further enhance our ability to improve oil quantity and quality. With the development of technologies in omics, it is now possible to observe the large amount of structural genomic variation, epigenetic states, alternative splicing of precursor mRNA, and variation of protein and metabolite, especially lipid metabolite, across natural inbred lines in maize. Discoveries of novel regulatory processes via these or other approaches will provide deep insight into the mechanisms underlying oil quantity and quality.

Acknowledgements We greatly appreciate Dr. Gen Xu in our laboratory for preparing two figures and reference list in this chapter. The funding is supported by the National Natural Foundation of China (31722039, 31361140362).

References

Alrefai R, Berke TG, Rocheford TR (1995) Quantitative trait locus analysis of fatty acid concentrations in maize. Genome 38(5):894–901

Baud S, Lepiniec L (2010) Physiological and developmental regulation of seed oil production. Prog Lipid Res 49(3):235–249

Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant J 50(5):825–838

Beisson F, Koo A, Ruuska S, Schwender J, Pollard M, Thelen J, Paddock T, Salas J, Savage L, Milcamps A (2003) Census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. Plant Physiol 132:681–697

Beló A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, Tingey S, Rafalski A (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Mol Genet Genomics 279(1):1–10

Benitez J, Gernat A, Murillo J, Araba M (1999) The use of high oil corn in broiler diets. Poult Sci 78(6):861–865

Cahoon RE, Heppard EP, Nagasawa N, Sakai H (2003) Alteration of embryo/endosperm size during seed development. US Patent 20,030,126,645

Chai Y, Hao X, Yang X, Allen WB, Li J, Yan J, Shen B, Li J (2011) Validation of DGAT1-2 polymorphisms associated with oil content and development of functional markers for molecular breeding of high-oil maize. Mol Breed 29(4):939–949

Clark D, Dudley JW, Rocheford TR, LeDeaux JR (2006) Genetic analysis of corn kernel chemical composition in the random mated 10 generation of the cross of generations 70 of IHO × ILO. Crop Sci 46(2):807

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158(2):824–834

Du H, Huang M, Hu J, Li J (2016) Modification of the fatty acid composition in *Arabidopsis* and maize seeds using a stearoyl-acyl carrier protein desaturase-1 (ZmSAD1) gene. BMC Plant Biol 16(1):137

Dudley JW (1977) Seventy-six generation of selection for oil and protein percentage in maize. In: Pollak E (ed) Proceedings of international conference on quantitative genetics. Iowa State University Press, Ames, pp 459–473

Dudley JW, Lambert RJ (2004) 100 generations of selection for oil and protein in corn. Plant Breed Rev 24:79–110

Dudley JW (2008) Epistatic interactions in crosses of Illinois High Oil × Illinois Low Oil and of Illinois High Protein × Illinois Low Protein corn strains. Crop Sci 48(1):59

Dudley JW, Clark D, Rocheford TR, LeDeaux JR (2007) Genetic analysis of corn kernel chemical composition in the random mated 7 generation of the cross of generations 70 of IHP × ILP. Crop Sci 47(1):45

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

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54(1):357–374

Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z, Wang B, Zhai L, Dai C, Xu J, Wang W, Li X, Zheng J, Chen L, Luo L, Liu J, Qian X, Yan J, Wang J, Wang G (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. Nat Commun 4:2832

Goldman IL, Rocheford TR, Dudley JW (1994) Molecular markers associated with maize kernel oil concentration in an Illinois high protein × Illinois low protein cross. Crop Sci 34(4):908–915

Han Y, Parsons C, Alexander D (1987) Nutritive value of high oil corn for poultry. Poult Sci 66(1):103–111

Han Y, Xu G, Du H, Hu J, Liu Z, Li H, Li J, Yang X (2017) Natural variations in stearoyl-acp desaturase genes affect the conversion of stearic to oleic acid in maize kernel. Theor Appl Genet 130(1):151–161

Hao X, Li X, Yang X, Li J (2014) Transferring a major QTL for oil content using marker-assisted backcrossing into an elite hybrid to increase the oil content in maize. Mol Breeding 34(2):739–748

Hopkins CG (1899) Improvement in the chemical composition of the corn kernel In: Dudley JW (ed) Seventy generation of selection for oil and protein in maize. ASA, CSSA, and SSSA, Madison, WI, pp 1–31

Jiao Y, Peluso P, Shi JH, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei W, Chin CS, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, Ware D (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527

Karn A, Gillman JD, Flint-Garcia SA (2017) Genetic analysis of teosinte alleles for kernel composition traits in maize. G3 7(4):1157–1164

Koh HJ, Heu MH, McCouch SR (1996) Molecular mapping of the ge s gene controlling the super-giant embryo character in rice (Oryza sativa L.). Theor Appl Genet 93(1):257–261

Lambert RJ, Alexander D, Mejaya I (2004) Single kernel selection for increased grain oil in maize synthetics and high-oil hybrid development. Plant Breed Rev 24:153–175

Lambert RJ (2001) High-oil corn hybrids. In: Hallau AR (ed) Special corn. CRC Press, Boca Raton, pp 131–153

Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B, Lai C, Clark D, Rocheford TR, Dudley JW (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168 (4):2141–2155

Leng ER (1962) Results of long term selection for chemical composition in maize and their significance in evaluating breeding systems In: Dudley JW (ed) Seventy generations of selection for oil and protein in Maize. ASA, CSSA, and SSSA, Madison, pp 149–173

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, DeBono A, Durrett TP, Franke RB, Graham LA, Katayama K, Kelly AA, Larson T, Markham JE, Miquel M, Molina I, Nishida I, Rowland O, Samuels L, Schmid KM, Wada H, Welti R, Xu C, Zallot R, Ohlrogge J (2010) Acyl-lipid metabolism. Arabidopsis Book 8(8):e0133

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y, Wang G, Li J, Yan J (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45(1):43–50

Li L, Li H, Li Q, Yang X, Zheng D, Warburton M, Chai Y, Zhang P, Guo Y, Yan J, Li J (2011) An 11-bp insertion in Zea mays fatb reduces the palmitic acid content of fatty acids in maize grain. PLoS ONE 6(9): e24699

Li Q, Yang H, Xu S, Cai Y, Zhang D, Han Y, Li L, Zhang Z, Gao S, Li J, Yan J (2012) Genome-wide association studies identified three independent polymorphisms associated with α-Tocopherol content in maize kernels. PLoS ONE 7(5):e36807

Lucas CJ, Zhao H, Schneerman M, Moose SP (2013) Genomic changes in response to 110 cycles of selection for seed protein and oil concentration in maize. In: Becraft PW (ed) Seed genomics. Wiley, pp 217–238

Môro GV, Santos MF, Bento DAV, Aguiar AM, de Souza CL (2012) Genetic analysis of kernel oil content in tropical maize with design III and QTL mapping. Euphytica 185(3):419–428

Mangolin C, De Souza C, Garcia A, Garcia A, Sibov S, de Souza A (2004) Mapping QTLs for kernel oil content in a tropical maize population. Euphytica 137 (2):251–259

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C (2009) Genetic properties of the maize nested association mapping population. Science 325(5941):737–740

Merlo AO, Cowen N, Delate T, Edington B, Folkerts O, Hopkins N, Lemeiux C, Skokut T, Smith K, Woosley A (1998) Ribozymes targeted to Stearoyl–ACP Δ9 Desaturase mRNA produce heritable increases of stearic acid in transgenic maize leaves. Plant Cell 10(10):1603–1621

Mikkilineni V, Rocheford TR (2003) Sequence variation and genomic organization of fatty acid desaturase-2 (fad2) and fatty acid desaturase-6 (fad6) cDNAs in maize. Theor Appl Genet 106(7):1326–1332

Miller R, Dudley J, Alexander D (1981) High intensity selection for percent oil in corn. Crop Sci 21(3):433–437

Mišević D, Alexander D (1989) Twenty-four cycles of phenotypic recurrent selection for percent oil in maize. I. Per se and test-cross performance. Crop Sci 29 (2):320–324

Moose SP, Dudley JW, Rocheford TR (2004) Maize selection passes the century mark: a unique resource for 21st century genomics. Trends Plant Sci 9(7):358–364

Moreno-Gonzalez J, Dudley J, Lambert R (1975) A design III study of linkage disequilibrium for percent oil in maize. Crop Sci 15(6):840–843

Ramsden CE, Zamora D, Leelarthaepin B, Majchrzak-Hong SF, Faurot KR, Suchindran CM, Ringel A, Davis JM, Hibbeln JR (2013) Use of dietary linoleic acid for secondary prevention of coronary heart disease and death: evaluation of recovered data

from the Sydney Diet Heart Study and updated meta-analysis. BMJ 346:e8707

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10(6):297–304

Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, Nubel D, Tarczynski MC (2010) Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant Physiol 153(3):980–987

Shen B, Roesler K (2017) Maize kenerl oil content In: Larkins BA (eds) Maize kernel development. Centre for Agriculture and Biosciences International, pp 160–174

Shintani DK, Ohlrogge JB (1994) The characterization of a mitochondrial acyl carrier protein isoform isolated from Arabidopsis thaliana. Plant Physiol 104(4):1221–1229

Song T, Chen S (2004) Long term selection for oil concentration in five maize populations. Maydica 49:9–14

Song X, Song T, Dai J, Rocheford TR, Li J (2004) QTL mapping of kernel oil concentration with high-oil maize by SSR markers. Maydica 49:41–48

Val LD, Schwartz SH, Kerns MR, Deikman J (2009) Development of a high oil trait for maize. In: Molecular genetic approaches to maize improvement. Springer, Berlin, pp 303–323

Wang H, Wu B, Song T, Chen S (2009) Effects of long-term selection for kernel oil concentration in KYHO, a high-oil maize population. Crop Sci 49 (2):459

Wang Z, Liu N, Ku L, Tian Z, Shi Y, Guo S, Su H, Zhang L, Ren Z, Li G, Wang X, Zhu Y, Qi J, Zhang X, Chen Y, Lübberstedt T (2016) Dissection of the genetic architecture for grain quality-related traits in three RIL populations of maize (*Zea mays* L.). Plant Breeding 135(1):38–46

Wassom JJ, Wong JC, Martinez E, King JJ, DeBaene J, Hotchkiss JR, Mikkilineni V, Bohn MO, Rocheford TR (2008a) QTL Associated with maize kernel oil, protein, and starch concentrations; kernel mass; and grain yield in Illinois high oil × B73 backcross-derived lines. Crop Sci 48(1):243

Wassom JJ, Mikkelineni V, Bohn MO, Rocheford TR (2008b) QTL for fatty acid composition of maize kernel oil in Illinois High Oil × B73 backcross-derived lines. Crop Sci 48(1):69

Weber EJ (1987) Lipids of the kernel. In: Watson SA, Ramstad PE (eds) Corn: chemistry and technology. American Association of Cereal Chemists, St. Paul, pp 311–349

Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017) Genome-wide association studies in maize: praise and stargaze. Mol Plant 10(3):359–374

Yan J, Warburton M, Crouch J (2011) Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Sci 51(2):433

Yang G, Li Y, Wang Q, Zhou Y, Zhou Q, Shen B, Zhang F, Liang X (2012) Detection and integration of quantitative trait loci for grain yield components and oil content in two connected recombinant inbred line populations of high-oil maize. Mol Breeding 29 (2):313–333

Yang X, Guo Y, Yan J, Zhang J, Song T, Rocheford TR, Li J (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. Theor Appl Genet 120 (3):665–678

Yang X, Ma H, Zhang P, Yan J, Guo Y, Song T, Li J (2012) Characterization of QTL for oil content in maize kernel. Theor Appl Genet 125(6):1169–1179

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178(1):539–551

Zhang H, Jin T, Huang Y, Chen J, Zhu L, Zhao Y, Guo J (2015) Identification of quantitative trait loci underlying the protein, oil and starch contents of maize in multiple environments. Euphytica 205(1):169–183

Zhang J, Lu X, Song X, Yan J, Song T, Dai J, Rocheford TR, Li J (2008) Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. Euphytica 162 (3):335–344

Zhang P, Allen WB, Nagasawa N, Ching AS, Heppard EP, Li H, Hao X, Li X, Yang X, Yan J, Nagato Y, Sakai H, Shen B, Li J (2012) A transposable element insertion within ZmGE2 gene is associated with increase in embryo to endosperm ratio in maize. Theor Appl Genet 125(7):1463–1471

Zheng P, Allen WB, Roesler K, Williams ME, Zhang S, Li J, Glassman K, Ranch J, Nubel D, Solawetz W, Bhattramakki D, Llaca V, Deschamps S, Zhong G, Tarczynski MC, Shen B (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat Genet 40(3):367–372

Zheng P, Babar MD, Parthasarathy S, Gibson R, Parliament K, Flook J, Patterson T, Friedemann P, Kumpatla S, Thompson S (2014) A truncated FatB resulting from a single nucleotide insertion is responsible for reducing saturated fatty acids in maize seed oil. Theor Appl Genet 127(7):1537–1547

Evolution and Adaptation in the Maize Genome

19

Nancy Manchanda, Samantha J. Snodgrass, Jeffrey Ross-Ibarra and Matthew B. Hufford

Abstract

Maize (*Zea mays* ssp. *mays*) has been a prime model organism for understanding the processes of domestication and adaptation. During domestication, maize underwent drastic morphological changes that differentiate it from its teosinte progenitor such as reduced tillering and seed shattering and freeing of the grain from a stony fruit case. Likewise, post-domestication adaptation to new environments has allowed maize to expand to a distribution far exceeding its wild relatives and in fact to a greater range than any other domesticate. Previous work using traditional top-down approaches, such as quantitative trait locus mapping and genome-wide association, has been successful in identifying canonical candidates for domestication and adaptation. However, the recent availability of genomic data and development of new analytical tools offer the opportunity to increasingly look at these processes from the bottom-up based on genomic signatures of selection. Here we review progress thus far in genomic research of maize domestication and adaptation. We discuss the insights genomics has shed on our understanding of these processes and conclude with a future outlook for how genomics might be further applied to these fields.

19.1 Introduction

Maize (*Zea mays* ssp. *mays*) serves as an excellent model organism for studying the genetic and functional mechanisms of evolution and adaptation. Over decades of research, resources including genome sequences, polymorphism (SNP) and expression data, gene model and functional annotations, mutant populations, gene-editing systems, mapping populations, and extensive germplasm collections have been generated and can be brought to bear on evolutionary questions.

While application of this wealth of resources would be promising in any system, it has been particularly fruitful in maize due to the crop's rich evolutionary history. Perhaps because of its

Nancy Manchanda and Samantha J. Snodgrass Equally Contributed.

N. Manchanda · S. J. Snodgrass · M. B. Hufford (✉)
Iowa State University, Ames, IA 50011, USA
e-mail: mhufford@iastate.edu

N. Manchanda
e-mail: nancym@iastate.edu

S. J. Snodgrass
e-mail: snodgras@iastate.edu

J. Ross-Ibarra
University of California, Davis, CA 95616, USA
e-mail: rossibarra@ucdavis.edu

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_19

high genetic diversity, maize has shown a remarkable capacity to adapt during domestication and subsequent colonization of diverse climates. For example, during domestication maize underwent a striking morphological transformation in both overall plant architecture and female inflorescence relative to its wild progenitor teosinte (*Zea mays* ssp. *parviglumis*). Similarly, maize experienced dramatic post-domestication adaptation to disparate environments including high elevation and temperate, long-day growing regions.

Over the last few decades, genomic technologies have been employed to unravel the mysteries of maize evolution. Researchers have sequenced and analyzed the DNA of extant wild and domesticated maize as well as ancient maize cobs from archeological sites. This work has offered a number of genomic insights into the early stages of maize domestication. Likewise, quantitative trait and association mapping studies have identified specific loci underlying traits relevant to domestication and adaptation such as inflorescence branching, ear shattering, flowering time, leaf angle, and plant pigmentation. Here we review advances in our understanding of both maize domestication and subsequent adaptation that have been made possible through the large-scale application of genomic data. We conclude by considering how innovations in genomic technology on the horizon will bring further clarity to our understanding of maize evolution.

19.2 Maize Domestication

Modern-day maize (*Zea mays* ssp. *mays*) was domesticated from wild *Zea mays* ssp. *parviglumis*, hereafter *parviglumis*, about 9,000 years ago in the Balsas River Basin of southwestern Mexico (Matsuoka et al. 2002; Piperno et al. 2009; van Heerwaarden et al. 2011). Because of the dramatic morphological differences between domesticated maize and *parviglumis*, for decades taxonomists debated the origin of maize, invoking scenarios involving extinct hybrid parents and backcrossing (Mangelsdorf 1947). Nonetheless, the combined evidence from crosses and cytology (Beadle 1972), archeology (Piperno et al. 2009), and genetics (Matsuoka et al. 2002; van Heerwaarden et al. 2011) clearly identified *parviglumis* as the direct wild progenitor. Recent work based on genome- and transcriptome-wide data has provided insight into the strength of the initial domestication bottleneck and additional aspects of maize demography and helped define the genetic architecture of natural selection during domestication.

19.2.1 The Domestication Bottleneck

The domestication process of maize and many other crops is known to include a demographic bottleneck in which only a small subset of the wild progenitor population contributes to the nascent domesticate. Such bottlenecks reduce effective population size and amplify the effects of genetic drift, leading to genome-wide reductions in diversity due to the stochastic loss of alleles. The domestication bottleneck also confounds efforts to identify targets of selection, as drift during the bottleneck can generate signals that can be mistaken for selection (Ross-Ibarra et al. 2007). Inference of adaptation from patterns of diversity in domesticated maize thus requires a clear understanding of demography.

The first efforts to estimate the bottleneck associated with maize domestication used coalescent simulations of sequence from the gene *Adh1* to estimate both the duration and the founding population size of the domesticated gene pool (Eyre-Walker et al. 1998). Eyre-Walker et al. (1998) estimated that maize underwent a severe bottleneck, with a maximum population size of 5,600 individuals over 2,800 years, or approximately 6% of the ancestral *parviglumis* population. Subsequent work is built upon this foundation, using 12 (Tenaillon et al. 2004) and then 774 loci (Wright et al. 2005) to obtain more accurate and detailed estimates. Most recently, Beissinger et al. (2016) took advantage of whole-genome sequence data from 23 maize and 13 teosinte lines (Chia et al. 2012) to model not only the population reduction

during the domestication bottleneck but also the process of rapid population expansion that occurred in maize post-domestication. Beissinger et al. (2016) estimated that genetic divergence between maize and *parviglumis* populations began about 15,000 generations before present and that the initial maize population was perhaps as small as 5% of *parviglumis*. Since then, maize has started to accumulate new diversity due to the mutational input facilitated by its dramatic expansion, with a current effective population size estimated to range from 370,000 to as high as 10 billion (Beissinger et al. 2016).

In addition to understanding the reduced diversity observed in maize relative to *parviglumis*, demography can also help explain the striking differences in deleterious alleles seen between maize and *parviglumis*. Wang et al. (2017) found that, while some deleterious alleles segregating in *parviglumis* were likely lost during the domestication bottleneck, a large number have become fixed in maize, resulting in a higher burden of deleterious load. Theoretical results predict exactly this kind of a response to deleterious recessive alleles under a demography including a population bottleneck (Simons et al. 2014). Wang et al. (2017) further found that this pattern was exacerbated during range expansion post-domestication due to serial founder effects, such that the load of deleterious alleles within maize lines is correlated with their geographic distance from the domestication center in the Balsas River Basin. Intriguingly, highland lines that have experienced introgression from the related wild teosinte *Zea mays* ssp. *mexicana* harbor fewer deleterious alleles in regions of introgression, consistent with the high historic effective size of the *mexicana* population (Ross-Ibarra et al. 2009) relative to maize, which results in more efficient purifying selection and removal of deleterious alleles. The understanding of maize demographic history that has been gained from recent genomic studies offers a clearer picture of ways in which diversity has been shaped by neutral processes over time and provides important context in the search for selected loci.

19.2.2 Identifying Domestication Loci

Since the 1970s, our understanding of the number of genes involved in the transition from wild to domesticated maize has changed considerably. Beadle (1972) estimated that as few as four or five genes controlled the substantial morphological differences between *parviglumis* and maize. Consistent with early inferences of a simple genetic architecture of domestication, two canonical domestication genes have been shown to have large, pleiotropic effects on maize morphology: *teosinte branched1* (*tb1*), involved in overall plant architecture and branching patterns, and *teosinte glume architecture1* (*tga1*), involved in the reduction of the stony fruit case that encapsulates teosinte seeds (Doebley 2004). However, subsequent studies based on full-genome data have shown hundreds if not thousands of genes and regulatory variants with signatures of selection during domestication (Hufford et al. 2012c; Wright et al. 2005) and also demonstrated that some domestication traits such as kernel row number, ear diameter, and tassel architecture have complex genetic architectures involving many genes of small effect (Lemmon and Doebley 2014; Xu et al. 2017).

Many have attempted to address this question using mapping approaches that work to associate a phenotype related to domestication (e.g, branching, shattering, dormancy) with a particular genotype. This "top-down" approach (see Fig. 19.1) has successfully identified a number of regions underlying domestication traits of interest. For example, Briggs et al. (2007) uncovered 59 QTL in a large backcross population between *parviglumis* and maize. Many of the QTL they identified, however, clustered in five genomic regions consistent with the large-effect loci identified from early crossing experiments. This approach is not without its challenges, however, including the well-documented phenomenon in which small-effect genes, when located in close proximity, behave as a single QTL, giving the false impression that one, large-effect gene controls the majority of a trait's variation (Beavis 1998). Such loci can be challenging to tease apart

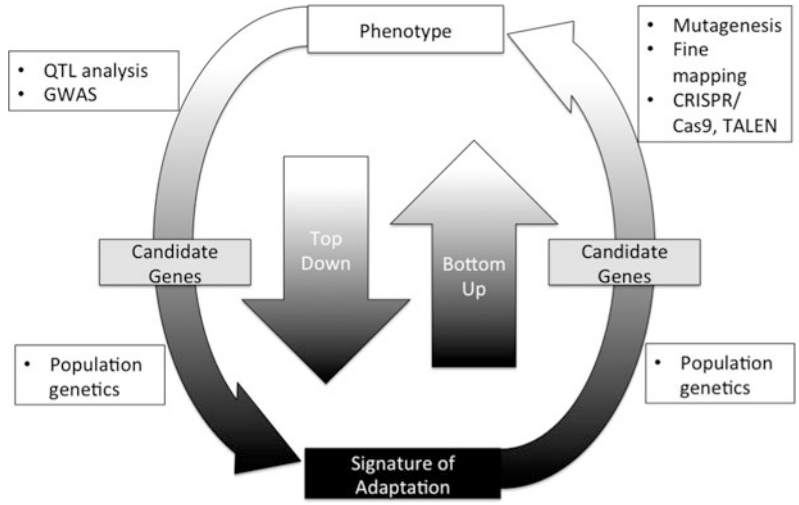

Fig. 19.1 Two approaches for studying domestication. A "top-down" approach begins with a phenotype and seeks to map and identify the causal loci. In contrast, a "bottom-up" method uses population genetic scans to detect signatures of selection and then reverse genetics approaches to identify function. Modified from Ross-Ibarra et al. (2007)

without the use of fine-scale mapping. Lemmon and Doebley (2014), for example, fine mapped a large-effect domestication locus on chromosome 5, fractionating this previously well-defined region into several smaller QTL affecting domestication traits and showing the underlying causal polymorphism was not in a single, pleiotropic gene.

In contrast to the top-down approach, a "bottom-up" method using population genetic scans to detect signatures of selection may have higher resolution in identifying loci and offers the potential to identify important selection candidates without a priori knowledge of the target phenotype (Ross-Ibarra et al. 2007). Early use of bottom-up approaches in maize scanned variation at microsatellite markers (Vigouroux et al. 2002) and in expressed sequence tag libraries (Wright et al. 2005) in order to detect selection. These methods sought loci showing reduced diversity in maize relative to *parviglumis* beyond the genome-wide expectation due to the domestication bottleneck. Such regions have likely experienced selection or are linked to a target of selection. Vigouroux et al. (2002) detected ten loci with strong signatures of selection that were in close proximity to previously identified domestication QTL. Likewise, Wright et al. (2005) estimated between 2 and 4% of the maize genome had experienced selection during domestication, translating to nearly 1,200 genes.

The drawback of these early efforts was the inability to survey loci across the genome, but with the advent of inexpensive next-generation sequencing, such constraints were quickly overcome. Hufford et al. (2012c) used over 21 million SNPs derived from whole-genome resequencing data from 35 improved maize lines, 23 traditional landraces, and 13 *parviglumis* lines for a comprehensive population genomic study of selection during domestication and subsequent crop improvement. Analyses of these data revealed three striking trends. First, focusing on the upper 10% of windows with the highest differentiation between *parviglumis* lines and landraces, Hufford et al. (2012c) found 484 chromosomal regions—roughly 8% of the maize genome, including more than 1,600 genes—potentially targeted by selection during domestication. Among the loci experiencing the most differentiation between maize and *parviglumis*, there were well-known loci including *tga1* and *tb1*, but also a large number of candidates with little to no known function within maize or no prior

evidence of being important for domestication. Hundreds of these unknown candidate regions showed stronger signals of selection than did *tga1* and *tb1*, highlighting the value of such a bottom-up population genomic approach. Putative functions for candidates were found through comparison to orthologs in other model systems and appear to be involved in pathways important for domestication such as flowering time (*zagl1*, GRMZM2G448355), phyllotaxy (*abph1*, GRMZM2G035688), and seed germination (GRMZM2G010290). Though subsequent work has validated important functional differences effected by a few of these candidate loci (e.g., (Sosso et al. 2015; Wills et al. 2018)), future efforts are warranted to tease apart their functional significance within maize.

While early population genetic models of selection focused on new beneficial mutations that quickly became fixed in a population (Smith and Haigh 1974), a second important finding from Hufford et al. (2012c) was that the majority of selection during domestication likely made use of alleles already segregating in *parviglumis* populations rather than new mutations: Of the more than 21 million SNPs studied, only ≈3,000 are fixed in maize and absent in *parviglumis*. The idea that existing variation in *parviglumis* could be selected to higher frequency in maize and result in a domesticated phenotype had also been proposed in earlier work by Lauter and Doebley (2002), who found "cryptic" genetic variation within *parviglumis* populations that could contribute to phenotypes only found in maize. In-depth molecular investigation at a number of loci (*gt1*, Wills et al. (2013); *tb1*, Studer et al. (2011); *ramosa1*, Sigmon and Vollbrecht (2010)) has identified the specific alleles selected during domestication and confirmed their presence in natural populations of *parviglumis*, sometimes at surprisingly high frequencies (*gt1*, Wills et al. (2013), *tb1*, Vann et al. (2015)). In fact, of all the domestication loci identified, only *tga1* is suspected to have arisen as a de novo mutation (Wang et al. 2015).

The third main result from the genome-wide analysis of Hufford et al. (2012c) was that intergenic regions contribute substantially to the genetic differences underlying maize domestication. Regions showing evidence of selection were on average ≈300 Kb in size, often making it difficult to identify the exact target of selection. Hufford et al. (2012c), however, found that domestication genes on average showed higher expression in maize than in *parviglumis*, suggesting much of the selection during domestication may have been on gene regulation. Consistent with this interpretation, Lemmon et al. (2014) used RNA-seq in maize-*parviglumis* F1 plants to show that genes harboring cis-regulatory differences were enriched for signals of selection, and Swanson-Wagner et al. (2012) found evidence of change across entire networks of gene regulation. Genome-wide, Hufford et al. (2012c) also found that 6% of candidate regions did not include any annotated gene sequences based on the B73 reference genome. While this serves as a lower bound on the proportion of selection that must have targeted gene regulation, subsequent efforts investigating the functional relevance of intergenic sequence have revealed that intergenic open chromatin is responsible for nearly one-third of the additive genetic variation across a number of phenotypic traits in maize (Rodgers-Melnick et al. 2016).

19.2.3 Domestication of Quantitative Traits

While population genetic analyses have primarily focused on individual genes, most phenotypic traits exhibit continuous phenotypic variation consistent with a highly polygenic genetic architecture. Recently, Xue et al. (2016), using two diverse panels of maize inbred lines, showed that many loci of small effect are responsible for extant variation in traits important for domestication. Previous work based on archeological evidence from a number of domesticated taxa has also suggested that evolution during domestication was in many cases rather gradual (Purugganan and Fuller 2011). Recent simulation efforts suggest that strong selection on a quantitative trait can lead to both gradual change in phenotype and evidence of selective sweeps at individual loci (Stetter et al. 2018).

Archeological evidence of maize cob size showed a gradual change (Benz et al. 2009), with cob diameter and cupule width changing only 1.25- and 1.83-fold over a period of 4,200 years. Also in line with expectations based on simulations, analysis of DNA from a 5,000-year-old cob revealed that the maize allele for the large-effect locus *tga1* was already present, but that the phenotype and genotype of the sample were intermediate between maize and *parviglumis* (Ramos-Madrigal et al. 2016). Their specimen likely represented an early step between *parviglumis* and maize and demonstrates that some genetic sweeps can occur rapidly even if phenotypes continue to gradually change over time, providing support for models including both large- and small-effect loci.

19.3 Maize Adaptation

Following initial domestication, maize rapidly spread from the Balsas River Basin in southwest Mexico to colonize both North and South America (Fig. 19.2), and, following the Columbian Exchange, now boasts the widest global extent of cultivation of any crop (Amato 1996; Rebourg et al. 2003; Merrill et al. 2009; Piperno 2011; Hake and Ross-Ibarra 2015). During this period of expansion, maize encountered a number of climates previously uninhabited by its progenitor, *parviglumis*. For example, while *parviglumis* populations are found at elevations of ≈800–1600 m and have a latitudinal range extending from approximately 16 to 22°N, maize can be found at elevations above 4000 m in the Andes and latitudes spanning 50°S to 50°N (Hufford et al. 2012a). Colonization of such diverse conditions required substantial adaptation in maize. In certain regions, this adaptation appears to have been achieved through gene flow from newly encountered, locally adapted populations of wild relatives (Hufford et al. 2013), whereas in others adaptation likely occurred de novo (Takuno et al. 2015). Thus far, investigation of genomic patterns of adaptation in both wild and cultivated maize has primarily considered selection across gradients of elevation and latitude.

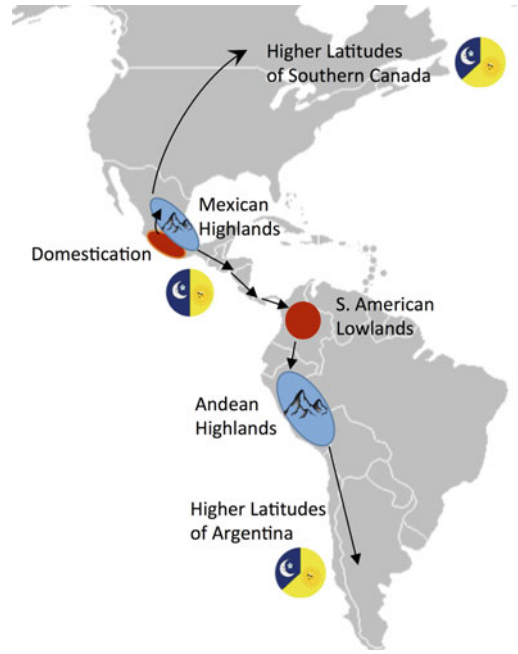

Fig. 19.2 Post-domestication spread of maize. Maize was domesticated in the Balsas River Valley of Mexico and then quickly spread throughout the Americas, colonizing high-elevation habitats in Central Mexico and the Andes from neighboring lowland regions (red ovals) and the high latitudes and long growing-season day lengths of what are now southern Canada and Argentina

19.3.1 What Can We Learn from Teosinte?

The genus *Zea* consists of five species (*Z. diploperennis*, *Z. luxurians*, *Z. mays*, *Z. nicaraguensis*, and *Z. perennis*) that are distributed from northern Mexico southward through Central America. Additionally, *Z. mays* includes four subspecies, the domesticate *Z. mays* ssp. *mays*, the aforementioned *parviglumis*, the highland-adapted *Z. mays* ssp. *mexicana* found in the Central Plateau of Mexico (hereafter, *mexicana*), and *Z. mays* ssp. *huehuetenangensis*, which is limited in its distribution to the midlands of northwest Guatemala. Collectively, wild taxa within *Zea* are referred to under the umbrella term "teosinte." Teosinte species diverged long before maize domestication (Ross-Ibarra et al. 2009) and are well adapted to their respective ecological niches (Hufford et al.

2012b). Maize is interfertile with all *Zea* species and, during its spread away from its domestication center in the Balsas River Basin, is thought to have received adaptive gene flow from newly encountered wild relatives (Hufford et al. 2013; Ross-Ibarra et al. 2009; Silva et al. 2015). An understanding of local adaptation in teosinte therefore informs our knowledge of both the historical adaptation process in maize and the potential resources at our disposal for adapting maize to future climates.

As the most widely distributed teosintes and the closest relatives of maize, *parviglumis* and *mexicana* are the only taxa for which detailed studies of local adaptation have been completed. Clear morphological differences distinguish the subspecies: Leaves and stems are green and glabrous (i.e., largely free from macrohairs) in *parviglumis* but deeply pigmented through anthocyanin accumulation and pilose (i.e., hairy) in *mexicana*. Pigment and pilosity are both thought to be adaptations of *mexicana* to the cooler high elevations of the Mexican Central Plateau. An early mapping experiment by Lauter et al. (2004), based on a population derived from a *parviglumis* X *mexicana* cross, helped dissect the genetic basis of differences in these traits, identifying several major and minor QTL.

Broad inferences regarding adaptation in *parviglumis* and *mexicana* have been made using environmental data from highly detailed species occurrence records and a machine learning approach which predicts a taxon's ecological niche (Hufford et al. 2012b). *Parviglumis* was found to occupy warm and thermally stable environments, whereas *mexicana* habitat was more variable and substantially cooler and drier. Very little overlap was observed between the modeled niches of the two subspecies, and their distributions were predicted to be stable since the Last Glacial Maximum (≈21,000 BP), suggesting *parviglumis* and *mexicana* are likely well adapted to distinct habitats (Hufford et al. 2012b).

One of the first large-scale surveys of genome-wide variation across multiple teosinte taxa identified putative adaptation through an inversion polymorphism (Fang et al. 2012). Theoretical work suggests that inversions may commonly play a role in adaptation by capturing locally adapted alleles and suppressing recombination between them (Kirkpatrick and Barton 2006), and empirical examples of adaptive inversions have been observed in both plant (Lowry and Wills 2010) and animal species (Hoffman et al. 2004; Kapun et al. 2016). Fang and co-authors identified a 50-Mb inversion (*Inv1n*) in teosinte on the short arm of chromosome 1 based on elevated linkage disequilibrium within this region. Additionally, through an analysis of haplotype clusters, the authors clarified that the derived, inverted haplotype was segregating in *parviglumis* and *mexicana*, but absent in other teosintes and domesticated maize, suggesting this structural polymorphism is rather new and likely only plays an adaptive role within these taxa. A strong clinal pattern was documented in the frequency of *Inv1n-I*, with decreasing prevalence at increasing elevation. This pattern, in combination with significant associations with phenotypic differences in culm diameter and tassel morphology, provided compelling evidence that the inversion plays an adaptive role in teosinte.

Subsequent studies of local adaptation in teosinte have improved the resolution of genetic markers and the genomic purview. For example, Pyhäjärvi et al. (2013) examined adaptive variation across 10 populations of *mexicana* and 11 populations of *parviglumis* using a data set of ≈37,000 SNPs. The authors identified considerable population structure and highly variable patterns in population-level diversity across the range of these subspecies. Additionally, based on patterns of linkage disequilibrium, the authors discovered three new inversion polymorphisms that were segregating within their sample, which, like the inversion polymorphism described by Fang et al. (2012), showed clinal allele frequency patterns across an elevation gradient consistent with local adaptation. Further candidate loci underlying adaptation were identified using four complementary approaches: analysis of allele frequency differentiation across populations, correlation between allele frequency and 76 environmental variables (soil, bioclimatic,

precipitation, altitude, etc.), estimation of adaptive variation based on each focal population, and, finally, an analysis of pairwise haplotype sharing. The authors identified numerous loci showing signatures of local adaptation across the species range based on these methods with candidate SNPs enriched in non-genic regions, revealing a potential role for regulatory variation in local adaptation. These results are in agreement with the proposal recently articulated by Mei et al. (2018), suggesting that genome size may play a role in how species adapt. For example, non-genic variation may be of more importance during local adaptation in large, highly repetitive genomes like maize, in stark contrast to compact genomes such as *Arabidopsis* where adaptation candidates are overwhelmingly genic (Hancock et al. 2011). Pyhäjärvi et al. (2013) further demonstrated that traits such as tassel morphology and flowering time showed signatures of selection across teosinte populations as did specific, compelling candidate genes such as *b1* in the anthocyanin pathway. Throughout this study, elevation played a clear and important role in structuring diversity and as a driver of local adaptation.

Recent genome-wide studies of local adaptation have more explicitly explored adaptation of *parviglumis* and *mexicana* across elevation gradients. Both Fustier et al. (2017) and Aguirre-Liguori et al. (2017) utilized samples collected across two elevational transects in the Mexican states of Jalisco and Guerrero. While Fustier et al. (2017) found reasonable overlap with adaptation candidates detected in Pyhäjärvi et al. (2013), showing, for example, selection on the *Inv1n* inversion, their analyses revealed novel patterns of adaptation including a soft selective sweep on chromosome 9 in a region known to underlie variation in macrohair content (Moose et al. 2004), as well as selection on traits linked to soil quality including root morphology, aluminum tolerance, and adaptation to low phosphorus availability. The key result emerging from Aguirre-Liguori et al. (2017), which surveyed a more extensive set of populations, was the detection of varying levels of local adaptation across the elevation gradient in western Mexico. The authors found that, as populations became more isolated from the center of their ecological niche going down- or upslope, the prevalence of loci showing signatures of adaptation increased, suggesting local adaptation becomes even more important at environmental extremes.

19.3.2 Maize Adaptation to High Elevation

One of the earliest migrations of maize was from the lowlands of the Balsas River Valley to the highlands of the Mexican Central Plateau. This human-assisted migration brought maize into sympatry with the highland teosinte *mexicana* for the first time. As described above, *mexicana* has multiple morphological features (e.g., high levels of pigmentation and pilosity) that are thought to be adaptive at high elevation. Maize from the highlands of Mexico shares these traits and is often found to hybridize with *mexicana* in the field, suggesting adaptive gene flow may have occurred between these taxa during colonization.

To test this hypothesis, Hufford et al. (2013) collected maize and *mexicana* from nine sympatric population pairs in the highlands of Central Mexico to evaluate the prevalence and directionality of gene flow between these taxa. Substantial and asymmetric gene flow was detected, primarily from *mexicana* into maize. Regions of *mexicana* introgression into maize were also largely conserved across maize populations in the Mexican highlands, suggesting an adaptive architecture of introgression. In multiple instances, introgressions from *mexicana* overlapped QTL controlling macrohairs and pigment (Lauter et al. 2004). A growth chamber experiment confirmed that maize with *mexicana* introgression showed higher fitness than maize lacking introgression under simulated high-elevation conditions.

As maize spread away from Mexico and outside of the distribution of its wild relatives, it colonized additional high-elevation regions (e.g., Guatemala, the southwestern USA, and the Andes) in which local wild populations of teosinte are absent. In colonizing and adapting to these

new highland regions, maize could have potentially carried highland haplotypes from Central Mexico when transiting lowland regions, or, alternatively, adapted de novo using the same or additional regions of the genome. To explore these possibilities, Takuno et al. (2015) assembled and generated genome-wide SNP data for a sample of 96 maize landraces from the lowlands and highlands of both Mexico and South America. After modeling and accounting for neutral demography in these samples, the authors assessed the genetic basis of high-elevation adaptation in both geographic regions, finding only limited evidence of convergent selection at the nucleotide or genic levels. This empirical evidence showing limited convergence was confirmed with a theoretical investigation of the likelihood that an allele that was adaptive in the highlands of Mexico could first survive transit of the lowlands of Central America and northern South America, where it was presumably deleterious, and then rise in frequency upon encountering highland conditions in the Andes. The likelihood of such a migration was exceedingly low.

Recently, Wang et al. (2017), using higher density, full-genome sequence data confirmed that *mexicana* haplotypes cannot be readily detected in the Andes. In contrast, these authors found substantial evidence of *mexicana* ancestry in maize landraces from the highlands of Guatemala and the southwestern USA, suggesting these populations share a genetic basis of high-elevation adaptation with maize in Mexico. The findings of Wang et al. (2017) are consistent with genetic analysis of archeological samples from the southwestern USA, which documented substantial *mexicana* ancestry in ancient maize in this region (da Fonseca et al. 2015).

19.3.3 Maize Adaptation to Temperate Latitudes

Maize migration to higher latitudes required substantial adaptation to both shorter growing season and differences in photoperiod. For example, whereas the growing season (i.e., frost-free days) in Mexico corresponds to the rainy period spanning roughly from May/June through October/November (≈180 days), the growing season in southern Canada is typically 130 days or less. Likewise, typical growing-season day lengths in Mexico are approximately 13 h, yet maize at higher latitudes is cultivated under 16-h day length conditions. While maize adaptation to novel temperate conditions is of interest from an evolutionary perspective, it has also received keen attention due to the importance of effective corn breeding for adaptation to various "relative maturity" zones. The development of large mapping populations that include both tropical and temperate maize has been critical for the identification of QTL and ultimately candidate genes underlying flowering time and photoperiod sensitivity.

In one of the most comprehensive studies to date, Buckler et al. (2009) utilized the maize nested association mapping (NAM) population for a joint QTL analysis. The NAM population contains 25 diverse inbred lines crossed to B73 (25 families), with 200 recombinant inbred lines (RILs) developed per family for a total of 5,000 lines (McMullen et al. 2009). In their analysis of flowering time using this resource, Buckler et al. (2009) identified 36 major QTLs for days to anthesis, 39 for days to silking, and 29 for anthesis–silking interval. However, the majority of QTLs had effect sizes that were small, most accounting for less than a day, which suggests the genetic architecture of flowering time in maize is complex and involves many additive, small-effect loci (Buckler et al. 2009). This result from maize is in stark contrast to other model organisms studied for flowering time—such as *Arabidopsis* (Alonso-Blanco et al. 1998), rice (Yano et al. 1997), barley (Turner et al. 2005), and sorghum (Lin et al. 1995)—where single, large-effect loci control most of the phenotypic variation. The authors suggested that mating system and demography potentially influence the overall genetic architecture of adaptive traits like flowering time (Buckler et al. 2009).

More recently, Romero Navarro et al. (2017) applied GWAS and a new method, F-one association mapping (FOAM), to a large and more diverse collection of maize landraces in order to

look for loci controlling flowering time. The authors considered the effects of both latitude and elevation on flowering time, finding 366 genes with significant association with elevation and 1,498 genes with significant association with latitude. Results from the FOAM study indicated that areas of low recombination in landraces such as inversion polymorphisms and centromeres had 20-fold enrichment for significant, previously undetected SNPs, suggesting these variants were purged during improvement of modern inbred lines. The researchers also identified 883 and 881 genes significantly associated with days to male and female flowering, respectively, with a small subset of loci (10 for male, 12 for female flowering) showing particularly strong associations (Romero Navarro et al. 2017).

In contrast to Buckler et al. (2009) and Romero Navarro et al. (2017) who assessed flowering time and temperate adaptation using large-scale mapping approaches in extant materials, Swarts et al. (2017) addressed this topic from the perspective of archeological genomics of maize from the southwestern USA. While maize was introduced into this region 4,000 years ago, it was not until 2,400–1,800 years ago that it became an agricultural mainstay. Swarts et al. (2017) used full-genome sequencing data from 1,844-year-old cobs sampled in a dry cave on the temperate plateau of Utah in combination with a crop model built using data from the Ames Inbred Diversity panel (Romay et al. 2013) to connect genotype and phenotype. Their predictive model was then validated on extant landraces from the southwestern USA as a proxy for the ancient sample lineage. From their model, Swarts et al. (2017) then predicted that the ancient samples would flower one week earlier on average than the lowland, desert-adapted sample from nearby sites (predictive accuracy was evaluated to be 0.72), resulting in a maturity cycle of 151 days in an environment with an average 149 frost-free days. Thus, these ancient samples were likely only marginally adapted to their environment (Swarts et al. 2017). Additionally, Swarts et al. (2017) estimated that temperate adaptation from subtropical material required approximately 2,000 years, which has substantial implications for breeding efforts for adaptive lines under rapid environmental change (Swarts et al. 2017).

In addition to the broadscale conclusions regarding the genetic architecture and evolution of flowering time and photoperiod sensitivity that have been made based on mapping populations like the maize NAM, a handful of genes underlying these traits have been identified using these resources. For example, functional and association studies by Ducrocq et al. (2008), Bouchet et al. (2013), and Meng et al. (2011) demonstrated the importance of *Vgt1* and *Vgt2*, respectively, during temperate adaptation. Similarly, through joint linkage analysis and GWAS followed by targeted high-resolution mapping, Hung et al. (2012) identified a candidate underlying photoperiod sensitivity in maize, *ZmCCT10* (i.e., *ZmCCT*), a homologue of the rice photoperiod response regulator *Ghd7*. Analysis of expression at *ZmCCT10* in F1 crosses between eight different teosinte and three temperate maize lines under long-day length conditions revealed higher expression of teosinte alleles when compared to maize alleles in all cases. This higher expression in teosinte is thought to repress expression of the florigen *ZCN8* (Meng et al. 2011) that is required to initiate flowering. Subsequently, Yang et al. (2013) and Huang et al. (2018) found evidence consistent with a CACTA-like transposon insertion at *ZmCCT10* playing a causal role in temperate adaptation. Huang et al. (2018) also identified a flowering time QTL on chromosome 9 near a second *CCT* transcription factor, (*ZmCCT9*). Analysis of expression patterns of maize and teosinte alleles at *ZmCCT9* under long-day conditions revealed that a Harbinger-like transposon acts in *cis* to regulate the expression of *ZmCCT9*, which then controls expression levels of the florigen *ZCN8* (Meng et al. 2011). Comparison of sequence from teosinte and tropical and temperate maize revealed that both the Harbinger-like transposon insertion at *ZmCCT9* and the CACTA-like transposon insertion at *ZmCCT10* were absent in teosinte and are likely

de novo mutations that occurred after initial maize domestication. The mutations then rose in frequency as maize colonized temperate regions.

19.4 Conclusions and Future Directions

Maize has indeed been a fruitful system for understanding domestication and adaptation. Using genomic data, researchers have been able to delve into the various evolutionary forces driving these processes. High-resolution models have given better estimates of the effects of demography on genetic diversity during domestication and range expansion. Further studies have demonstrated the importance of standing variation, convergent evolution, and adaptive gene flow during the history of maize. Moreover, adaptive inversions, regulatory mutations, and intergenic sequence variation all have demonstrable impacts on phenotypic variation and have likely aided adaptation to novel environments such as temperate latitudes and highlands.

Given the trove of insight achieved thus far, continual development of genomic and analytical methods promises the capacity to address what have previously been unfeasible questions. For instance, previous research has shown the importance of standing variation and polygenic traits as targets of selection during domestication and adaptation. Yet these processes were difficult to study with existing methods and greater resolution may be obtained through recent innovations in machine learning.

Notable gaps in genomic data will also need to be filled to facilitate evolutionary study. For example, while gene flow between maize and *mexicana* is relatively well documented and understood, our knowledge of gene flow to and from other teosinte species such as *Z. diploperennis* is in its infancy, awaiting the generation of additional population-level data for these taxa. Furthermore, with the generation of multiple de novo whole-genome assemblies, scientists could use comparative genomic methods to analyze the diversity of structural variants and their relevance to adaptation, including large and small inversions, copy number variation, and presence–absence variation.

While genomic methods have revealed much about the maize genome, they have also opened the door to entirely new questions. The answers to these questions will not only be important for our basic knowledge of plant biology and evolution, but could potentially inform ongoing improvement of maize. Identification and characterization of both adaptive gene flow from diverse *Zea* taxa and climatic associations in maize could help pinpoint alleles for breeding that are beneficial in certain environments. Similarly, generation of several de novo genome assemblies will broaden our understanding of the link between genotype and phenotype and allow for a pan-genome approach to maize improvement in which structural and copy number variation could be selected intentionally. Rapidly improving methods for annotation of transposable elements, open chromatin, and DNA methylation will provide a more complete understanding of how this variation has played a role in maize evolution and adaptation and expand the breeding toolkit further still. In short, genomic data have substantially improved our understanding of evolution and the maize genome, and, given emerging technologies and remaining unanswered questions, the future of this field is promising indeed.

References

Aguirre-Liguori JA, Tenaillon MI, Vazquez-Lobo A, Gaut BS, Jaramillo-Correa JP, Montes-Hernandez S, Souza V, Eguiarte LE (2017) Connecting genomic patterns of local adaptation and niche suitability in teosintes. Mol Ecol 26(16):4226–4240

Alonso-Blanco C, El-Assal SED, Coupland G, Koornneef M (1998) Analysis of natural allelic variation at flowering time loci in the landsberg erecta and cape verde islands ecotypes of arabidopsis thaliana. Genetics 149(2):749–764

Amato J, Hudson JC (1996) Making the corn belt: a geographical history of middle-western agriculture (midwestern history and culture.). Am His Rev 101 (2):576

Beadle GW (1972) The mystery of maize. Field Mus. Bull. 212–221

Beavis WD (1998) QTL analyses: power, precision, and accuracy. CRC Press, pp 145–162

Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J (2016) Recent demography drives changes in linked selection across the maize genome. Nat Plants 2(16):084

Benz B, Cheng L, Leavitt SW, Eastoe C (2009) El riego and early maize agricultural evolution. Left Coast Press, Inc., Chap 5

Bouchet S, Servin B, Bertin P, Madur D, Combes V, Dumas F, Brunel D, Laborde J, Charcosset A, Nicolas S (2013) Adaptation of maize to temperate climates: mid-density genome-wide association genetics and diversity patterns reveal key genomic regions, with a major contribution of the vgt2 (zcn8) locus. PLoS ONE 8(8):1–17

Briggs WH, McMullen MD, Gaut BS, Doebley J (2007) Linkage mapping of domestication loci in a large maize teosinte backcross resouce. Genetics 177:1915–1928

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Sofia da Silva H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325 (5941):714–718

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC et al (2012) Maize hapmap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807

Doebley J (2004) The genetics of maize evolution. Annu Rev Genet 38(1):37–59

Ducrocq S, Madur D, Veyrieras JB, Camus-Kulandaivelu L, Kloiber-Maitz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178(4):2433–2437

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci 95 (8):4441–4446

Fang Z, Pyhäjärvi T, Weber AL, Dawe RK, Glaubitz JC, González JdJS, Ross-Ibarra C, Doebley J, Morrell PL, Ross-Ibarra J (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. Genetics 191 (3):883–894

da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe C, Ávila-Arcos MC, Hufnagel DE, Korneliussen TS, Vieira FG, Jakobsson M, Arriaza B, Willerslev E, Nielsen R, Hufford MB, Albrechtsen A, Ross-Ibarra J, Gilbert MTP (2015) The origin and evolution of maize in the Southwestern United States. Nat Plants 1. http://dx.doi.org/10.1038/nplants.2014.3

Fustier M, Brandenburg J, Boitard S, Lapeyronnie J, Eguiarte LE, Vigouroux Y, Manicacci D, Tenaillon MI (2017) Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. Mol Ecol 26(10):2738–2756

Hake S, Ross-Ibarra J (2015) The natural history of model organisms: genetic, evolutionary and plant breeding insights from the domestication of maize. eLife https://doi.org/10.7554/elife.05861

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J (2011) Adaptation to climate across the arabidopsis thaliana genome. Science 334(6052):83–86

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci 108(3):1088–1092

Hoffmann AA, Sgrò CM, Weeks AR (2004) Chromosomal inversion polymorphisms and adaptation. Trends Ecol Evol 19(9):482–488

Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, Xu G, Tian J, Wang C, Li D, Wu L, Yang X, Jin W, Doebley JF, Tian F (2018) ZmCCT9 enhances maize adaptation to higher latitudes. Proc Natl Acad Sci 115 (2):E334–E341

Hufford MB, Bilinski P, Pyhäjärvi T, Ross-Ibarra J (2012a) Teosinte as a model system for population and ecological genomics. Trends Genet 28(12):606–615

Hufford MB, Martinez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI (2012b) Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. PLOS ONE 7 (11):1–9

Hufford MB, Xu X, van Heerwaarden J, Pyhäjä T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012c) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811

Hufford MB, Lubinksy P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J (2013) The genomic signature of crop-wild introgression in maize. PLoS Genet 9(e1003):477

Hung HY, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF, Holland JB (2012) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc. Natl Acad Sci 109(28):11,068–11,069

Kapun M, Fabian DK, Goudet J, Flatt T (2016) Genomic evidence for adaptive inversion clines in drosophila melanogaster. Mol Biol Evol 33(5):1317–1336

Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. Genetics 173(1):419–434

Lauter N, Doebley J (2002) Genetic variation for phenotypically invariant traits detected in teosinte: Implications for the evolution of novel forms. Genetics 160(1):333–342

Lauter N, Gustus C, Westerbergh A, Doebley J (2004) The inheritance and evolution of leaf pigmentation and pubescence in teosinte. Genetics 167(4):1949–1959

Lemmon ZH, Doebley JF (2014) Genetic dissection of a genomic region with pleiotropic effects on domestication traits in maize reveals multiple linked QTL. Genetics 198:345–353

Lemmon ZH, Bukowski R, Sun Q, Doebley JF (2014) The role of cis regulatory evolution in maize domestication. PLoS Genet 10(11):1–15

Lin YR, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the poaceae, in reference to an interspecific sorghum population. Genetics 141(1):391–411

Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol 8(9):1–14

Mangelsdorf PC (1947) The origin and evolution of maize. Adv Genet 1:161–2017

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci 99(9):6080–6084

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Rosas MO, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. Science 325(5941):737–740

Mei W, Stetter MG, Gates DJ, Stitzer MC, Ross-Ibarra J (2018) Adaptation in plant genomes: bigger is different. Am J Bot 105(1)

Meng X, Muszynski MG, Danilevskaya ON (2011) The ft-like zcn8 gene functions as a floral activator and is involved in photoperiod sensitivity in maize. Plant Cell 23(3):942–960

Merrill WL, Hard RJ, Mabry JB, Fritz GJ, Adams KR, Roney JR, MacWilliams AC (2009) The diffusion of maize to the Southwestern United States and its impact. Proc Natl Acad Sci 106(50):21,019–21,026

Moose SP, Lauter N, Carlson SR (2004) The maize macrohairless1 locus specifically promotes leaf blade macrohair initiation and responds to factors regulating leaf identity. Genetics 166(3):1451–1461

Piperno DR (2011) The origins of plant cultivation and domestication in the new world tropics 52(S4):S453–S470

Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium b.p. maize from the central Balsas river valley, Mexico. Proc Natl Acad Sci 106 (13):5019–5024

Purugganan MD, Fuller DQ (2011) Archaeological data reveal slow rates of evolution during plant domestication. Evolution 65(1):171–183

Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. Genome Biol Evol 5(9):1594–1609

Ramos-Madrigal J, Smith BD, Moreno-Mayar JV, Gopalakrishnan S, Ross-Ibarra J, Gilbert MTP, Wales N (2016) Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. Curr Biol 26:3195–3201

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106(5):895–903

Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES (2016) Open chromatin reveals the functional maize genome. Proc Natl Acad Sci

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES, Gardner CA (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14(6):R55

Romero Navarro JA, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, Ortega A, Banda AE, Montiel NOG, Ortiz-Monasterio I, Vincente FS, Espinoza AG, Atlin G, Wenzl P, Hearne S, Buckler ES (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet 49:476–480

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc Natl Acad Sci 104(suppl 1):8641–8648

Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus zea. Genetics 181(4):1399–1413

Sigmon B, Vollbrecht E (2010) Evidence of selection at the ramosa1 locus during maize domestication. Mol Ecol 19(7):1296–1311

Silva NCdA, Vidal R, Costa FM, Vaio M, Ogliari JB (2015) Presence of *Zea luxurians* (durieu and ascherson) bird in Southern Brazil: implications for the conservation of wild relatives of maize. PLoS ONE 10 (10):1–16

Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. Nat Genet 46:220–224

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23(1):23–35

Sosso D, Luo D, Li QB, Sasse J, Yang J, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS, Rogowsky PM, Ross-Ibarra J, Yang B, Frommer WB (2015) Seed filling in domesticated maize and rice depends on sweet-mediated hexose transport. Nat Genet 47:1489–1493

Stetter MG, Thornton K, Ross-Ibarra J (2018) Genetic architecture and selective sweeps after polygenic

adaptation to distant trait optima. BioRxiv 313247. https://doi.org/10.1101/313247

Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet 43:1160–1163

Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM (2012) Reshaping of the maize transcriptome by domestication. Proc Natl Acad Sci 109(29):11,878–11,883

Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peeples M, Lepak N, Prim L, Romay MC, Ross-Ibarra J, Sanchez-Gonzalez JdJ, Schmidt C, Schuenemann VJ, Krause J, Matson RG, Weigel D, Buckler ES, Burbano HA (2017) Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science 357 (6350):512–515

Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, Hufford MB, Ross-Ibarra J (2015) Independent molecular basis of convergent highland adaptation in maize. Genetics 200:1297–1312

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 21(7):1214–1225

Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator ppd-h1 provides adaptation to photoperiod in barley. Science 310 (5750):1031–1034

Vann L, Kono T, Pyhäjärvi T, Hufford MB, Ross-Ibarra J (2015) Natural variation in teosinte at the domestication locus *teosinte branched1* (*tb1*). PeerJ 3:e900

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci 99 (15):9650–9655

Wang H, Studer AJ, Zhao Q, Meeley R, Doebley JF (2015) Evidence that the origin of naked kernels during maize domestication was caused by a single amino acid substitution in TGA1. Genetics 200 (3):965–974

Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford MB (2017) Genome Biol 18

Wills DM, Whipple CJ, Takuno S, Kursel LE, Shannon LM, Ross-Ibarra J, Doebley JF (2013) From many, one: genetic control of prolificacy during maize domestication. PLoS Genet 9(6):1–13

Wills DM, Fang Z, York AM, Holland JB, Doebley JF (2018) Defining the role of the mads-box gene, zea agamous-like1, a target of selection during maize domestication. J Hered 109(3):333–338

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308 (5726):1310–1314

Xu G, Wang X, Huang C, Xu D, Li D, Tian J, Chen Q, Wang C, Liang Y, Wu Y, Yang X, Tian F (2017) Complex genetic architecture underlies maize tassel domestication. New Phytol 214(2):852–864

Xue S, Bradbury P, Casstevens T, Holland JB (2016) Genetic architecture of domestication-related traits in maize. Genetics 204:13–99

Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T, Li J, Chen Y, Yan J, Yang X, Xu M (2013) Cacta-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc Natl Acad Sci 110(42):16,969–16,974

Yano M, Harushima Y, Nagamura Y, Kurata N, Minobe Y, Sasaki T (1997) Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. Theor Appl Genet 95 (7):1025–1032

Part V

Application: Allele Mining and Genomics-Assisted Breeding

Harnessing Maize Biodiversity

20

Luis Fernando Samayoa, Jeffrey C. Dunne, Ryan J. Andres and James B. Holland

Abstract

The phenotypic and genetic diversity of maize worldwide is remarkable. This chapter summarizes decades of studies of the genetic diversity of maize populations from different parts of the world, methods used to group maize into informal hierarchies, and how these groupings partition genetic and trait variation. The USA is the most important maize-producing nation, but the genetic diversity of USA maize is small relative to the available worldwide variation. Tropical maize harbors more genetic variation, but is not adapted to growing in temperate environments. Two distinct approaches to tapping the global reservoir of maize diversity to improve USA and other temperate region maize crops are outlined. One approach, allele mining, involves discovery of alleles with large favorable effects on traits in exotic germplasm, followed by marker-aided backcrossing or gene editing to introduce specific unique alleles into elite breeding populations. Alternatively, for traits conditioned mostly by many small-effect polygenes, rapid genomic selection for adaptation followed by combining ability within pure exotic populations could be used to create adapted and improved versions of exotic populations before they are crossed to elite adapted inbreds to make new breeding populations.

20.1 Genomic and Phenotypic Diversity in Maize

Maize is a remarkably diverse species, adapted to a wide range of climatic conditions and farming practices. The latitudinal range of maize is immense, ranging from 54°N in Alberta, Canada, to 45°S in the province of Chubut, Argentina (Monfreda et al. 2008) (http://www.earthstat.org/data-download/). In terms of altitude, maize is cultivated from sea level to 4000 m (Confite Puneño in Peru) (Grobman et al. 1961). In Mexico alone, maize is grown in arid regions receiving 400 mm of rain to tropical environments with 3555 mm of precipitation during their growing season (Ruiz Corral et al. 2008).

It is widely accepted that maize was domesticated from the wild grass teosinte (*Z. mays*

The original version of this chapter was revised: For detailed information please see correction. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-97427-9_22

L. F. Samayoa · J. C. Dunne
Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695-7620, USA

R. J. Andres · J. B. Holland (✉)
United States Department of Agriculture – Agriculture Research Service, Plant Science Research Unit, Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695-7620, USA
e-mail: Jim.Holland@ARS.USDA.GOV

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes, https://doi.org/10.1007/978-3-319-97427-9_20

ssp. *parviglumis*) in southwestern Mexico approximately nine thousand years ago (Matsuoka et al. 2002b; Piperno and Flannery 2001). Maize was spread by human migration and exchanged widely throughout the Americas over the next few thousand years and was already grown from Canada to Chile prior to the arrival of Columbus (Weatherwax 1954; Crawford et al. 2006). As part of the Columbian exchange five hundred years ago, maize was taken to Europe and subsequently spread to Africa and Asia (Rebourg et al. 2003; Crosby 1972). As a result, maize was exposed to a plethora of environmental conditions and agro-ecosystems, resulting in the evolution of diverse adaptations by a combination of natural and human selection (Aguilar et al. 2003; Ruiz de Galarreta and Alvarez 2001; Hung et al. 2012). In addition, humans selected maize for a wide variety of food and ceremonial purposes (Hernández Xolocotzi 1972, 1985; Logan et al. 2012; Ortega Paczka 2003). Countering the diversifying selection for distinct adaptations and uses was the exchange of different seed types within and among cultural groups and the outcrossing nature of maize, which promoted significant gene flow among maize varieties. Similar complex processes continue today in rural Mexico (Pressoir and Berthaud 2004a, b; Hugo et al. 2009; Brush and Perales 2007; Bellon 1991; Rubey et al. 1997), although there is evidence that the genetic diversity of maize is eroding in Mexico due to socioeconomic pressures (Dyer et al. 2014; Ortega Corona et al. 2013).

The first attempt to organize the enormous diversity of maize was made by Sturtevant (1894), who proposed six groups of maize based mostly on endosperm type. Anderson and Cutler (1942) proposed a natural classification (better reflecting phylogenetic relationships) of maize into races. Anderson and Cutler's (1942) broad definition of race was 'a group of related individuals with enough characters in common to permit their recognition as a group.' The necessarily imprecise definition of racial groups in maize reflects a common problem in biology: What is the best way to group living organisms that vary continuously without clear delineations? Races represent a level of organization below subspecies (as all cultivated maize are classified as *Zea mays* L. subsp. *mays* (Iltis and Doebley 1980)) but above 'variety.' Maize plants vary continuously, both phenotypically and genotypically, such that any system of grouping requires imposing arbitrary boundaries based on similarity or relatedness. A hallmark of natural classifications is they reflect patterns of evolutionary descent, and so are hierarchical: 'the patterned complexity of living systems is hierarchically organized' (Mayr 1982). Thus, within maize, various levels of classification are possible, from racial complexes (Bird and Goodman 1977) to races, to accessions and collections (which often represent a population of maize grown at one location at one time), to individual plants (Fig. 20.1).

Maize races are defined based primarily on ear, seed, tassel, and whole plant morphology along with geographic distribution (Goodman and Brown 1988). Following Sturtevant (1894), the next effort to describe the variation in North American maize was by Chávez (1913), who considered the different maize types from Mexico and the USA to be varieties and also proposed a system of grouping based on kernel type. Girola (1919) published a similar evaluation of maize types grown in Argentina, which included some varieties of Peruvian and North American origin. The first major effort to collect samples of maize diversity in Latin America appears to have been led by S. M. Bukasov of the Soviet Institute of Applied Botany in 1925–27 as part of Nikolai Vavilov's efforts to study and understand the diversity and utility of crops from around the world (Kuleshov 1930). Kuleshov (1933) evaluated more than 8000 populations of maize sampled globally, finding a huge range in number of leaves per plant (9–43 among different populations) and plant height (60–700 cm). The fast maturing varieties from extreme latitudes were short-statured with few leaves, whereas the slow-growing maize from some tropical regions in the Yucatan and southern Colombia produced tall plants with many vegetative nodes. The greatest phenotypic diversity was concentrated in Mexico and the Andes (Kuleshov 1933).

Detailed characterization of the variation in Mexican maize was done by Anderson (1946)

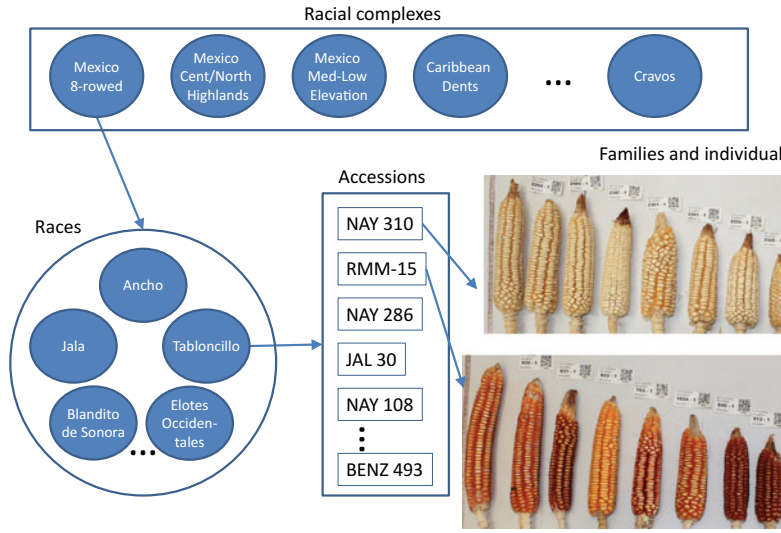

Fig. 20.1 Hierarchical distribution of phenotypic diversity in maize. A subset of Mexican races are shown here as an example

and more extensively by Wellhausen et al. (1951). The study by Wellhausen et al. (1951) involved classification of maize into races and detailed attempts to interpret the origin and relationships among different races. Wellhausen et al. described the challenge this way:

> Frequently there are no sharp lines of demarcation between the varieties or races which compromise a cultivated species or genus…Since maize not only belongs to a single species but is also largely cross-fertilized, it offers more than the ordinary number of difficulties to the taxonomist. Hence, it is not surprising that the classification of maize, in spite of its importance, should have been so long neglected. Taxonomists who shun cultivated plants as not botanically important may actually be avoiding difficult problems not easily solved by traditional taxonomic methods. The variation in cultivated plants is frequently so bewildering that additional techniques including those of the geneticist, the cytologist, and the agronomist are needed to bring a semblance of order out of apparent chaos… the classification of maize presented…has made use not only of the morphological characteristics of the ear, the tassel and the plant, but also of genetic, cytological, physiological and agronomic characteristics. Special consideration has been given to geographical distribution.

Following the achievement of collecting and classifying much of the Mexican maize diversity, the National Academy of Sciences' National Research Council sponsored an extensive project to collect, characterize, and preserve the variation of maize throughout all of Latin America. Different teams of scientists assembled collections for a country or region, classified maize types into races, and published monographs on their regional maize (scanned versions of all of the monographs are available at https://www.ars.usda.gov/midwest-area/ames/plant-introduction-research/docs/races-of-maize/). Lacking an objective criterion for declaring populations to be of the same or different races, racial assignments were made by careful, if subjective, grouping based mostly on ear and seed morphology. Ramirez et al. (1960) described the methods for classifying Bolivian maize as follows:

> Ears were laid out on long tables…collections which looked similar were placed near each other on the same table, paying attention to color, texture, and size of grain; number of rows and size and shape of ear. It was immediately apparent that several of the races previously described…were present in the collections. The remaining races were worked out by distinctive combinations of form and color…Work was continued until a preliminary classification had been worked out using only the appearance of the ears themselves. The notes on location and altitudes of the original collections were then consulted. This on the whole confirmed previous judgments but in one case it demonstrated that two somewhat similar races had been confused and in others that certain classifications made largely on slight color differences had ignored general similarities in form. When the racial composition of the varieties was provisionally worked out, as many

> selections from each race as possible were planted for study…the photographs of the typical ears were then carefully compared with the plants grown from them…In this way it was possible to follow what George Box calls the iterative process of discovery. Hypotheses based on a study of the ears were checked by an examination of the plants. Hypotheses based on plant study could be confirmed by examining the ears. On the basis of these observations new and more inclusive hypotheses could be formed. When a combination of characters which tended to go together was found it could be recognized and used in working out the racial composition of the varieties.

Subsequent work was undertaken to classify European and Asian maize as well; those monographs are also available at https://www.ars.usda.gov/midwest-area/ames/plant-introduction-research/docs/races-of-maize/. Additional classification work has been performed for the Great Lakes region of North America (Azar et al. 1997), Europe (Gouesnard et al. 1997; Ruiz de Galarreta and Alvarez 2001), Asia (Kumar et al. 2015), and Africa (Ndiso et al. 2013). The races of Mexico are perhaps the best studied, and since many populations are still extant and maintained by farmers in Mexico to this day, new collections have been made throughout the entire country in the last decade, allowing researchers to study the changes in preference for maize races since the original collections (Ortega Corona et al. 2013). Approximately 350 races of maize have been recognized in the Americas alone (Goodman and Brown 1988; Vigouroux et al. 2008). A distinguishing feature of most races is that they were historically preserved as local open-pollinated varieties (landraces), with only individual plant selection for local adaptation and end use characteristics. Breeding methods such as progeny testing, replicated family tests, inbreeding, and deliberate cross-breeding were never practiced on the vast majority of maize landrace populations. Only a few of the 350 or more races worldwide have contributed in a significant way to modern inbred lines and hybrid cultivars (Goodman 2005).

The authors of the race monographs recognized the subjective nature of the racial assignments and considered them to be preliminary classifications requiring subsequent re-evaluation. Important questions that remained following these initial classifications were: (1) How much variation exists within and among races? (2) How different are races from each other? (3) How are races from different regions related? (4) Do racial classifications based on key morphological traits accurately reflect phylogenetic (or overall genomic) relationships? (5) Is 'functional' genetic variation (i.e., sequence variations that affect phenotypes) distributed among maize populations in the same way as neutral genetic variation? (6) How much molecular variation exists beyond SNPs: transposable elements, repetitive DNA, structural variation, and genome size variation? Some answers to these questions have been obtained through DNA marker analysis of landrace and other diverse germplasm collections and will be discussed in detail below.

20.1.1 How Much Variation Exists Within and Among Races?

Molecular variability in maize landraces was first examined with isozyme marker systems. There is an exceptional level of variation of these enzymes across a worldwide collection of landraces (Doebley et al. 1983, 1984, 1985; Goodman and Stuber 1983a, b; Pflüger and Schlatter 1996; Sanou et al. 1997; Sánchez et al. 2000a, b) (Table 20.1), and *Zea* (maize and teosinte included) contains more isozyme variation than all other plant species studied (Doebley et al. 1984). A series of papers described the isozyme diversity and relationships among maize races within specific countries or regions of Latin America and parts of the USA (Goodman and Stuber 1983a; Goodman and Brown 1988; Sánchez et al. 2000a; Doebley et al. 1983, 1986; Bretting et al. 1987, 1990; Sánchez and Goodman 1992a, b; Sánchez et al. 2000a; Sánchez et al. 2006, 2007). Most of these studies reported allelic richness (number of alleles per locus) and expected heterozygosity as measures of genetic diversity within populations (Table 20.1). Substantial variation was observed in the genetic variation within populations, and many rare

Table 20.1 Estimates of genetic variability of maize landraces at local country, continental and global levels using isozymes or DNA-based markers

Maize races studied				Genetic markers		Genetic diversity			Alleles/locus (mean)		References
Origin	No. plants per accession	No. accessions	No. races	No	Type	Within accessions	Within races	Total among and within races	Within races	Total among and within races	
Americas											
Bolivia	12	101	31	23	Isozymes				1.83	5.17	Goodman and Stuber (1983a)
Southwestern USA	12	45		21	Isozymes					4.48	Doebley et al. (1983)
Americas and Caribbean	12	72	15	23	Isozymes					4.83	Bretting et al. (1987)
Guatemala	12	67	19	23	Isozymes					6.65	Bretting et al. (1990)
Americas	**12**	***1080***	***390***[a]	**23**	**Isozymes**		***0.19***		***2.0***	***14.3***	**Sánchez et al.** (2000b)
Argentina	12	28	22	23	Isozymes		0.17		1.7	3.7	Sánchez et al. (2000b)
Bolivia	12	101	31	23	Isozymes		0.16		1.8	5.0	Sánchez et al. (2000b)
Brazil	12	61	29	23	Isozymes		0.19		1.8	3.7	Sánchez et al. (2000b)
Chile	12	49	20	23	Isozymes		0.18		1.8	4.1	Sánchez et al. (2000b)
Colombia	12	88	23	23	Isozymes		0.19		2.1	6.0	Sánchez et al. (2000b)
Ecuador	12	118	34	23	Isozymes		0.19		2.0	6.3	Sánchez et al. (2000b)
Guatemala	12	66	19	23	Isozymes		0.23		2.5	6.5	Sánchez et al. (2000b)
Mexico	12	209	59	23	Isozymes		0.23		2.6	9.0	Sánchez et al. (2000b)
Peru	12	141	48	23	Isozymes		0.16		1.8	6.0	Sánchez et al. (2000b)
USA	12	135	77[a]	23	Isozymes		0.19		1.8	6.1	Sánchez et al. (2000b)
Venezuela	12	65	19	23	Isozymes		0.19		2.1	4.6	Sánchez et al. (2000b)
W. Indies	12	19	8	23	Isozymes		0.17		1.7	3.0	Sánchez et al. (2000b)
Brazil and adjacent areas	12	193	89	23	Isozymes		0.19			7.0	Sánchez et al. (2007)
Andean region	12	439	131	23	Isozymes		0.17			9.9	Sánchez et al. (2006)
Mexico	12	209	59	37	Isozymes		0.24	0.27		8.2	Sánchez et al. (2000a)
Mexico	12	93	34	23	Isozymes		0.25			7.09	Doebley et al. (1985)
Mexico	1	161	46	37 k	SNPs		0.31				Arteaga et al. (2016)
Latin America	?	28	28	15	cpSSRs			0.37		2.80	Provan et al. (1999)
Argentina	8–30	19	12	12	SSRs		0.47		3.79	10.5	Bracco et al. (2012)
Latin America	**6**	***4***	**4**	**42**	**SSRs**			***0.61***		***5.60***	Matsuoka et al. (2002a)

(continued)

Table 20.1 (continued)

Maize races studied				Genetic markers		Genetic diversity			Alleles/locus (mean)		References
Origin	No. plants per accession	No. accessions	No. races	No	Type	Within accessions	Within races	Total among and within races	Within races	Total among and within races	
Mexico	6	2	2	42	SSRs		0.50		3.03		Matsuoka et al. (2002a)
Venezuela	6	2	2	42	SSRs		0.51		2.93		Matsuoka et al. (2002a)
Mexico	13–21	25	24	25	SSRs		0.48	0.61	3.44	7.84	Reif et al. (2006)
Mexico	16–21	23	23	25	SSRs		0.48	0.61		7.84	Warburton et al. (2008)
Mexico	20	31	1	2	cpSSRs	0.49					Pressoir and Berthaud (2004a)
Mexico	20	31	1	11	SSRs	0.71					Pressoir and Berthaud (2004a)
Mexico	25	107	9	31	SSRs		0.50		11.2		Vega-Alvarez et al. (2017)
Americas	***1–2***	***404***	**238**	**96**	**SSRs**			***0.83***		***39.0***	**Vigouroux et al.** (2008)
Andean	1–2	170	98	96	SSRs			0.71		12.4	Vigouroux et al. (2008)
Northern USA	1–2	29	26	96	SSRs			0.72		10.6	Vigouroux et al. (2008)
Tropical lowland	1–2	133	76	96	SSRs			0.80		14.4	Vigouroux et al. (2008)
Highland Mexican	1–2	72	38	96	SSRs			0.81		14.9	Vigouroux et al. (2008)
Latin America	***15***	***194***	***131***	**28**	**SSRs**			***0.62***		***10.4***	**Bedoya et al.** (2017)
South America—Andean region	15	65	61	28	SSRs			0.57		8.07	Bedoya et al. (2017)
Mesoamerican lowland	15	58	37	28	SSRs			0.59		7.89	Bedoya et al. (2017)
Mexico and Southern Andes	15	46	30	28	SSRs			0.63		8.10	Bedoya et al. (2017)
Mixed populations	15	25	24	28	SSRs			0.61		7.21	Bedoya et al. (2017)
Europe											
Europe	**2 × 15 bulk**	***488***		**23**	**RFLPs**			***0.51***	***2.48***	***11.5***	**Gauthier et al.** (2002)
East	2 × 15 bulk	45		23	RFLPs				2.04	6.52	Gauthier et al. (2002)
Balkans	2 × 15 bulk	17		23	RFLPs				2.36	5.74	Gauthier et al. (2002)
Central France	2 × 15 bulk	47		23	RFLPs				2.50	6.83	Gauthier et al. (2002)
Italy	2 × 15 bulk	106		23	RFLPs				2.45	7.61	Gauthier et al. (2002)
Greece	2 × 15 bulk	50		23	RFLPs				2.17	6.61	Gauthier et al. (2002)
Pyrénées	2 × 15 bulk	50		23	RFLPs				2.71	7.61	Gauthier et al. (2002)
Spain	2 × 15 bulk	102		23	RFLPs				2.55	8.04	Gauthier et al. (2002)

(continued)

Table 20.1 (continued)

Maize races studied				Genetic markers		Genetic diversity			Alleles/locus (mean)		References
Origin	No. plants per accession	No. accessions	No. races	No	Type	Within accessions	Within races	Total among and within races	Within races	Total among and within races	
Portugal	2 × 15 bulk	71		23	RFLPs				2.75	7.48	Gauthier et al. (2002)
Europe	**2 × 15 bulk**	***131***		**29**	**RFLPs**			***0.55***	***2.49***	***9.59***	**Rebourg et al.** (2001)
Northeastern	2 × 15 bulk	32		29	RFLPs			0.50	2.02	6.2	Rebourg et al. (2001)
Southeastern	2 × 15 bulk	14		29	RFLPs			0.54	2.53	5.7	Rebourg et al. (2001)
Italy	2 × 15 bulk	16		29	RFLPs			0.49	2.3	5.5	Rebourg et al. (2001)
France	2 × 15 bulk	32		29	RFLPs			0.51	2.72	6.7	Rebourg et al. (2001)
Southwestern	2 × 15 bulk	37		29	RFLPs			0.55	2.78	7.9	Rebourg et al. (2001)
Europe	2 × 15 bulk	404		19	Isozymes					3.1	Revilla et al. (2003)
Europe	30	5		55	SSRs	0.41		0.53	3.3	5.9	Reif et al. (2005)
Other regions											
Africa	15	20		20	SSRs			0.50		7.3	Oppong et al. (2014)
China	15	54		42	SSRs			0.69		7.86	Yao et al. (2007)
China	15	124		45	SSRs			0.70		7.93	Qi-Lun et al. (2008)
India	15	48		42	SSRs			0.63		13.0	Sharma et al. (2010)
Intercontinental											
Americas (1115)	***1***	***1115***	**351**	**259**	**SNPs**			**0.27**			**Westengen et al.** (2012) **and van Heerwaarden et al.** (2011)
West and Highland Mexico	1	177		259	SNPs			0.28			Westengen et al. (2012) and van Heerwaarden et al. (2011)
Mesoamerican lowland/Coastal Brazil	1	284		259	SNPs			0.25			Westengen et al. (2012) and van Heerwaarden et al. (2011)
Lowland South America	1	173		259	SNPs			0.23			Westengen et al. (2012) and van Heerwaarden et al. (2011)
Andean highland/Bolivian lowland	1	266		259	SNPs			0.19			Westengen et al. (2012) and van Heerwaarden et al. (2011)
Northern America	1	215		259	SNPs			0.27			Westengen et al. (2012) and van Heerwaarden et al. (2011)

(continued)

Table 20.1 (continued)

Maize races studied				Genetic markers		Genetic diversity			Alleles/locus (mean)		References
Origin	No. plants per accession	No. accessions	No. races	No	Type	Within accessions	Within races	Total among and within races	Within races	Total among and within races	
Africa	***1***	***48***		**259**	**SNPs**			**0.25**			**Westengen et al.** (2012)
Eastern	1	30		259	SNPs			0.25			Westengen et al. (2012)
Western	1	6		259	SNPs			0.18			Westengen et al. (2012)
Sahelian	1	12		259	SNPs			0.24			Westengen et al. (2012)
Europe	2 × 15 bulk	131		24	SSRs	0.38				7.8	Dubreuil et al. (2006)
Americas	2 × 15 bulk	144		24	SSRs	0.44				7.8	Dubreuil et al. (2006)
Americas + Europe	**2 × 15 bulk**	***217***		**29**	**RFLPs**	***0.38***		***0.58***		***13.1***	**Rebourg et al.** (2003)
Europe	**2 × 15 bulk**	***129***		**29**	**RFLPs**	***0.35***		***0.55***		***9.55***	**Rebourg et al.** (2003)
Northeastern Europe	2 × 15 bulk	30		29	RFLPs	0.27		0.48		6.10	Rebourg et al. (2003)
Southeastern Europe	2 × 15 bulk	14		29	RFLPs	0.37		0.54		5.72	Rebourg et al. (2003)
Italy	2 × 15 bulk	16		29	RFLPs	0.33		0.49		5.45	Rebourg et al. (2003)
France	2 × 15 bulk	32		29	RFLPs	0.39		0.51		6.66	Rebourg et al. (2003)
Southwestern Europe	2 × 15 bulk	37		29	RFLPs	0.41		0.55		7.90	Rebourg et al. (2003)
Americas	**2 × 15 bulk**	***88***		**29**	**RFLPs**	***0.42***		***0.59***		***12.34***	**Rebourg et al.** (2003)
Northern America	2 × 15 bulk	31		29	RFLPs	0.43		0.59		8.62	Rebourg et al. (2003)
Central America	2 × 15 bulk	19		29	RFLPs	0.44		0.57		9.62	Rebourg et al. (2003)
Caribbean Islands	2 × 15 bulk	20		29	RFLPs	0.37		0.47		6.66	Rebourg et al. (2003)
North Andean region	2 × 15 bulk	5		29	RFLPs	0.37		0.50		5.28	Rebourg et al. (2003)
Southern South America	2 × 15 bulk	13		29	RFLPs	0.43		0.55		6.93	Rebourg et al. (2003)
Americas	1 × 15 bulk	258		17	SSRs	0.44		0.60		9.00	Mir et al. (2013)
Asia	1 × 15 bulk	143		17	SSRs	0.38		0.55		6.94	Mir et al. (2013)
Africa/Middle East	1 × 15 bulk	250		17	SSRs	0.39		0.57		6.88	Mir et al. (2013)
Europe	1 × 15 bulk	148		17	SSRs	0.41		0.59		6.47	Mir et al. (2013)

Genetic diversity values were reported as expected heterozygosity (H_e) or a closely related statistic in most publications. For studies encompassing large geographic regions, the overall diversity statistics are reported in bold font, followed by subregion statistics in normal font

1 × 15 bulk—bulk of DNA from 15 plants, 2 × 15 bulk—2 bulks of DNA from 15 plants each, *cpSSRs*—chloroplast simple sequence repeats, *No.*—number, *RFLPs*—restriction fragment length polymorphisms, *SNPs*—single-nucleotide polymorphisms, *SSRs*—simple sequence repeats

[a]For the purposes of this table, USA and Canada varieties were considered races. A comprehensive racial classification of maize of the USA and Canada has not been performed

alleles were observed, often occurring only in one accession. The highest proportion of unique alleles was found in Guatemala and Mexico, surrounding the most likely center of maize domestication, while the Andes represented another region of exceptionally high diversity (Sánchez et al. 2000b). Higher levels of diversity were found in geographically widespread races, such as Tuxpeño, Conico, Chalqueño, Hickory King, Oloton, and San Marceño, which have had importance as sources of commercial corn varieties (Sánchez et al. 2000b). Races wither lower values of intrapopulation diversity were mostly populations used for special food uses, such as pozole, popcorn, sweet corn, and blue corn. Their lower diversity was attributed to small effective population sizes due to restricted areas planted to these types of maize (Sánchez et al. 2000a, b).

More recent evaluations of genetic diversity in maize have used DNA markers, starting with restriction fragment length polymorphisms (RFLPs), followed by simple sequence repeats (SSRs), and more recently single-nucleotide polymorphisms (SNPs) and other sequence-based methods. One pattern that is readily noticeable in Tables 20.1 and 20.2 is that marker system has a strong influence on genetic diversity estimates. For example, all the estimates of genetic diversity across widely different sets of populations based on isozymes, RFLPs, or SSRs, indicate genetic diversity levels above 0.5 and an average of at least 7 alleles per locus (Tables 20.1 and 20.2). In contrast, the more recent SNP estimates all report values considerably lower than 0.5, and the vast majority of SNPs have only two alleles reported (a maximum of only four variants per site is possible). Why do SNPs indicate much lower levels of genetic diversity? Diversity estimates from SNPs tend to be lower than SSRs because the smaller number of alleles per locus and the strong shift toward rare allele frequencies in the site frequency spectrum reduces their discriminatory value, as observed when the two different marker systems are used on the same genetic samples (Jones et al. 2007; Hamblin et al. 2007; Van Inghelandt et al. 2010). It is important, therefore, to compare genetic diversity across different samples only on the basis of a common set of markers, or at least a common marker type.

For the purposes of maize breeding, comparisons of genetic diversity within commercial hybrid cultivars or breeding lines (Table 20.2) to global maize germplasm (Table 20.1) are important. These comparisons indicate the proportion of genetic diversity that is captured in cultivars and breeding programs; however, they are complicated by the sampling procedures use to choose representative germplasm. Maize landraces contain 83% of the sequence variation found in the wild ancestor *Zea mays* subsp. *parviglumis* (Vigouroux et al. 2005), whereas maize inbreds sampled globally have been estimated to contain from 77 to 98% of the sequence variation found in landraces (Tenaillon et al. 2001; Hufford et al. 2012; Liu et al. 2003). The striking variation among these latter estimates is hard to understand; the relatively small germplasm samples in these studies contribute to uncertainty in the estimates. The highest estimate of variation retained in breeding lines is based on a comparison to inbreds extracted from landraces (Hufford et al. 2012), so some selection against deleterious recessive alleles and alleles conferring tropical adaptation likely occurred, reducing the variation in the landrace samples. Much higher estimates of reduced diversity in temperate inbreds and hybrids were estimated from isozyme studies. For example, Smith et al. (1985) compared isozyme variation in a sample of 72 widely used inbred lines from the USA to previously reported variation within Bolivian races (Goodman and Stuber 1983a), finding that 'fifty plants from a single Bolivian race were found to have approximately the same number of isozyme alleles as the 72 US lines.' These values indicate that 'exotic' inbred lines contain many distinct alleles not represented in temperate maize breeding programs and maize landraces contain even more alleles not captured in any inbred sets.

Table 20.2 Estimates of genetic variability of maize inbreds, hybrids, and open-pollinated varieties

Germplasm studied				DNA markers		Genetic diversity		Alleles per locus (mean)	References
Origin	No. plants per accession	No. accessions	Population type	No	Type	Within accessions	Total	Total	
Global tropical/subtropical	1	58	Inbred lines	94	SSRs		0.81	13.5	Liu et al. (2003)
Sweet corn	1	5	Inbred lines	94	SSRs		0.64	2.9	Liu et al. (2003)
Temperate non-Stiff Stalk	1	94	Inbred lines	94	SSRs		0.78	12.8	Liu et al. (2003)
Popcorn	1	7	Inbred lines	94	SSRs		0.54	2.9	Liu et al. (2003)
Temperate Stiff Stalk	1	33	Inbred lines	94	SSRs		0.59	5.7	Liu et al. (2003)
Mixed	1	63	Inbred lines	94	SSRs		0.61	14.1	Liu et al. (2003)
USA	10	40	Inbred lines	83	SSRs		0.67	4.9	Lu and Bernardo (2001)
USA	1	92	Ex-PVP commercial inbred lines	768	SNPs		0.35	1.9	Nelson et al. (2008)
USA	1	13	Public inbred lines		SNPs		0.33	1.9	Nelson et al. (2008)
Mexico	~22	23	Landraces	25	SSR	0.48	0.61	7.8	Warburton et al. (2008)
CIMMYT	~29	23	OPV	25	SSR	0.54	0.61	8.5	Warburton et al. (2008)
CIMMYT	1	261	Inbred lines	25	SSR		0.65	8.5	Warburton et al. (2008)
Global	***1***	***102***	***Inbred lines***	**42**	**SSRs**		***0.62***	***6.9***	Matsuoka et al. (2002a)
Global tropical	1	41	Inbred lines		SSRs		0.59	5.5	Matsuoka et al. (2002a)
USA	1	54	Inbred lines		SSRs		0.60	5.2	Matsuoka et al. (2002a)

(continued)

Table 20.2 (continued)

Germplasm studied				DNA markers		Genetic diversity		Alleles per locus (mean)	References
Origin	No. plants per accession	No. accessions	Population type	No	Type	Within accessions	Total	Total	
Canada/Europe	1	7	Inbred lines		SSRs		0.58	2.8	Matsuoka et al. (2002a)
Global	***1***	***279***	***Inbred lines***	**259**	**SNPs**		***0.25***	**–**	**Westengen et al. (2012)**
Global tropical/subtropical	1	83	TS	259	SNPs		0.24	–	Westengen et al. (2012)
Temperate Stiff Stalk	1	164	NSS	259	SNPs		0.25	–	Westengen et al. (2012)
Temperate non-Stiff Stalk	1	32	SS	259	SNPs		0.15	–	Westengen et al. (2012)

Genetic diversity values were reported as expected heterozygosity (H_e) or a closely related statistic in most publications. For studies encompassing large geographic regions, the overall diversity statistics are reported in bold font, followed by subregion statistics in normal font

No.—number, *SNPs*—single-nucleotide polymorphisms, *SSRs*—simple sequence repeats

20.1.2 How Different Are Races from Each Other?

Differentiation or 'structure' among subgroups or subpopulations within a species is typically quantified with F_{ST} or related measures (Weir and Cockerham 1984; Holsinger and Weir 2009). F_{ST} can be thought of as 'the proportion of genetic diversity due to allele frequency differences among populations' or, equivalently, the 'correlation of randomly chosen alleles within the same subpopulation relative to the entire population' (Holsinger and Weir 2009). Groups or subpopulations that have similar allele frequencies have low differentiation and low values of F_{ST}; F_{ST} increases as groups have greater differences in allele frequencies. A wide range of F_{ST} estimates has been reported for different samples of maize races and other groupings (Table 20.3). Differentiation among races based on isozymes or SSRs ranged from 0.08 to 0.34 among studies, compared to an average of 0.23 for differentiation among populations based on isozyme studies across a range of outcrossing crops (Hamrick and Godt 1997). Based on this comparison, the level of population differentiation observed among maize races is not remarkable; the diversifying selection on maize races is counterbalanced by gene flow from outcrossing, and migration aided by human seed exchange. The level of differentiation among races of maize is about equal to F_{ST} estimates among populations within its wild ancestor, *Zea mays* subsp. *parviglumis* (0.21) and greater than F_{ST} estimated between maize and *parviglumis*, which tends to be around 0.10 (Table 20.3).

The concept of differentiation as the proportion of variation due to differences among groups can be extended to multiple levels of nested hierarchical groupings (Holsinger and Weir 2009). A few estimates are available to compare differentiation among races (F_{RT}) and differentiation among accessions (typically collections from a single field) within races (F_{AR}). Note that differentiation at lower levels of the hierarchy (e.g., individuals or accessions) is at least equal to differentiation at higher levels (accessions or races); for example, the differentiation between individuals sampled from different accessions of different races depends upon variation among accessions within races, plus variation among races. F_{RT} was nearly as large as F_{AR} (0.10 vs. 0.13, respectively) in a survey of Mexican landraces (Sánchez et al. 2000a) and about half as large as F_{AR} (0.12 vs. 0.25, respectively) in Latin American and Caribbean landraces (Bedoya et al. 2017). These results suggest that racial classifications, as well as specific farmer collections within races, both represent truly distinct genetic groupings in most cases. However, race name alone is often not highly predictive of genetic similarity, in contrast to geographic origin information alone (Vigouroux et al. 2008).

A clear outlier value in Table 20.3 is the F_{ST} between landraces and global maize inbreds estimated based on sequence data as only 0.02 by Hufford et al. (2012). This value is surprisingly low and compares to estimates of F_{ST} ranging from 0.06 to 0.18 between tropical inbreds and two different groups of temperate inbreds by Liu et al. (2003). It is not clear why the estimate of differentiation between inbreds and landraces reported by Hufford et al. (2012) is so small, although this is the same study that estimated very little loss of diversity in inbreds compared to landraces and the caveats already expressed apply to estimates of F_{ST} in this sample as well. In contrast, the level of differentiation among modern USA heterotic groups estimated by van Heerwaarden et al. (2012) is the highest among all samples reported in Table 20.3. This extreme divergence between modern heterotic groups ('Stiff Stalk' and 'non-Stiff Stalk' temperate inbred pools) is all the more striking because the ancestral populations from which the modern lines were derived had little differentiation (F_{ST} = 0.05) (van Heerwaarden et al. 2012). Genetic drift and many cycles of selection for complementary sets of alleles that maximize yield performance when hybrids are made across groups are the mechanisms that lead to the creation of modern heterotic groups (Tracy and Chandler 2008). Although the authors suggest that ascertainment bias was likely small because genetic distances were highly correlated for different sets of markers, it is possible that F_{ST}

Table 20.3 F_{ST} values for differentiation among different taxa and levels of hierarchy of maize populations

Taxa	Markers	Fst (differentiation among groups)	References
Mexican races	Isozymes	0.27	Doebley et al. (1985)
Mexican races	SSRs	0.21	Reif et al. (2006)
Mexican races	Isozymes	0.10[a]	Sánchez G. et al. (2000)
Mexican accessions/races	Isozymes	0.21[a]	Sánchez G. et al. (2000)
Bolivian races	Isozymes	0.34[a]	Goodman and Stuber (1983a)
N. & S. American accessions/Races	Isozymes	0.15	Sánchez G. et al. (2000b)
Populations/outcrossing spp.	Isozymes	0.23	Hamrick and Godt (1997)
Z. parviglumis pops	SSRs	0.21	Hufford (2010)
Landraces versus parviglumis	21M SNPs	0.11	Hufford et al. (2012)
Landraces versus parviglumis	SSRs	0.07	Vigouroux et al. (2005)
Maize versus parviglumis	Gene sequences	0.08	Ross-Ibarra et al. (2009)
Mexican races	SSRs	0.21	Warburton et al. (2008)
N. & S. American races	SSRs	0.08	Vigouroux et al. (2008)
Latin America and Caribbean races	SSRs	0.12a	Bedoya et al. (2017)
Latin America and Caribbean accessions within races	SSRs	0.25a	Bedoya et al. (2017)
Argentine and Bolivian races	960 SNPs	0.27	Jamann et al. (2017)
Tropical versus SS versus NSS temperate inbreds	SSRs	0.06–0.18	Liu et al. (2003)
Global inbreds versus landraces	21M SNPs	0.02	Hufford et al. (2012)
Modern USA heterotic groups	50k SNPs	0.38	van Heerwaarden et al. (2012)

[a]These values are not reported in publication, but were computed by J. B. Holland from published original data

estimates have an upward bias in this study because many of the array-based markers were pre-selected to reveal differences between inbreds from the two modern USA heterotic groups. In this regard, sequence-based SNPs are not likely to be affected by ascertainment bias, but when comparing diverse maize sample sequences, alignment to the reference B73 inbred sequence may underrepresent the variation in more diverse materials, since poorly aligning sequences that have strongly diverged are likely to be excluded from analysis. This 'alignment bias' perhaps contributed to the very small F_{ST} estimate reported by Hufford et al. (2012).

20.1.3 How Are Races from Different Regions Related?

The original classifications of maize into races were performed on a regional basis, mostly independent of each other. Some comments were made in the original race monographs suggesting some relationships between races in different countries, but systematic analysis of groupings across political borders was not conducted until the studies of Goodman and Bird (Bird and Goodman 1977; Goodman and Bird 1977). These studies used numerical taxonomy of the original morphological data from the race

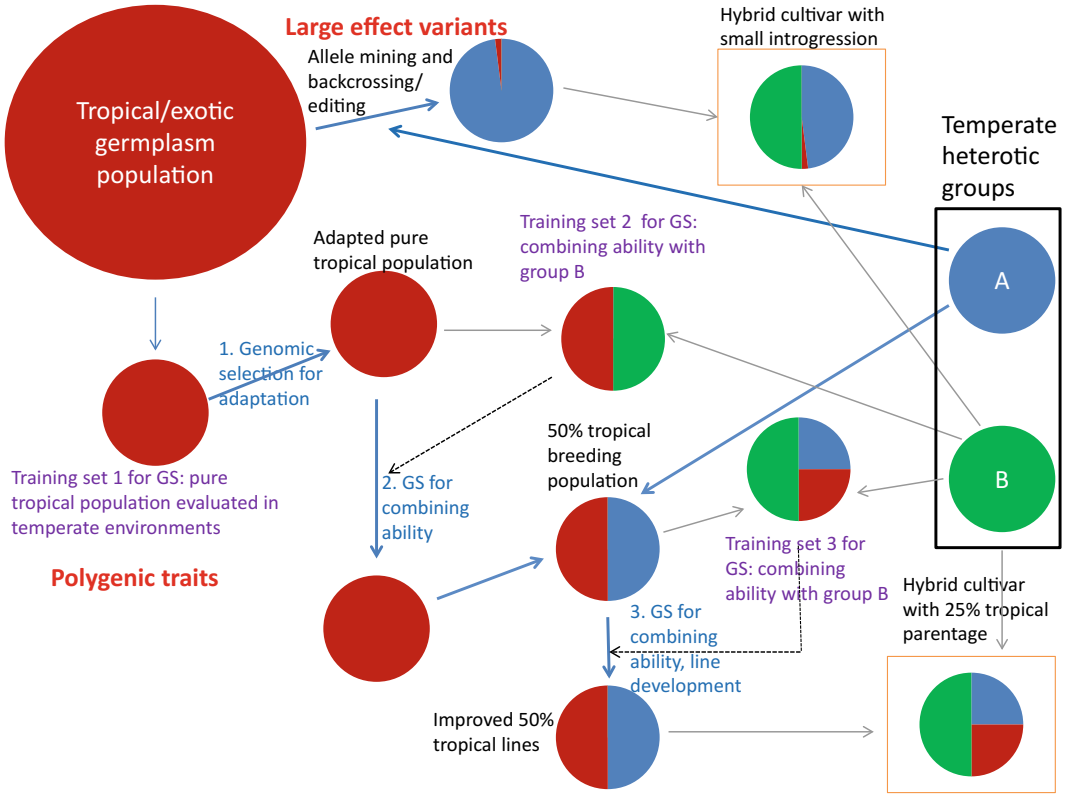

Fig. 20.2 Schematic workflow for improving temperate maize populations using alleles from unadapted tropical populations. Large effect 'single-gene' variants for important traits can be backcrossed or gene edited into temperate inbreds using allele mining techniques. Traits controlled by many genes each with small effects ('polygenic traits') require a more complex breeding plan. Here we propose three phases of genomic selection: 1. adaptation mainly by selection for earlier flowering time, 2. selection within pure tropical populations for combining ability with temperate testers, and 3. selection within mixed temperate/tropical populations for combining ability with temperate testers from a different heterotic group. Dark blue lines represent germplasm contributions to breeding populations. Thin solid lines represent crosses to form hybrids for evaluation or commercialization. Dashed lines represent phenotypic data used as input to genomic selection models. Orange boxes surround hybrid cultivars, resulting from the breeding program

monographs to estimate the relationships of races from different regions. This led to the identification of 'racial complexes'—a higher order grouping of related races (Fig. 20.1), summarized in Goodman and Brown (1988), who illustrated 61 large-scale racial complexes for maize in the Americas. These groupings have been revised and refined as additional data on genetic relationships among maize populations has been obtained.

Subsequent studies with SSRs added to and refined the morphology-based groupings of races. The most comprehensive analysis of SSR diversity in races of maize in the Americas was performed by Vigouroux et al. (2008). They found that the races of maize in the western hemisphere can be organized into four major groups: northern USA, highland Mexican, tropical lowland, and Andean races. Some groups of races can be described as admixtures or intermediate evolutionary steps between these four major groups. For example, southeastern USA races seem to have arisen from admixture of northern USA and tropical lowland groups. Southwestern USA races are intermediate between highland Mexican and northern USA,

but probably represent intermediate steps in the evolution of northern USA types from Mexican populations rather than admixture between the two.

An independent analysis of diversity of maize in the Americas was conducted by Bedoya et al. (2017), who confirmed the large-scale patterns suggested in the previous study. These two large-scale surveys of SSR variation in New World maize provide insight into the likely direction of maize dispersal throughout the Western Hemisphere. Based on genetic similarities, it appears that maize spread southward from its center of origin through Central American into northern South America, and then into the Andes and the Caribbean. Bedoya et al. (2017) suggested that Caribbean maize was influenced by migration from both northern South America and the Yucatan Peninsula, although Vigouroux et al. (2008) did not infer evidence for the Yucatan connection. Similarly, the two studies were not congruent with respect to the relationship between Andean maize and highland Guatemalan maize. Southern and central Brazilian maize was likely an admixture of Andean and tropical lowland maize, the latter introduced through Venezuela via the east coast of South America (Vigouroux et al. 2008). These studies also note the difficulty of inferring migration patterns in the face of complex historical movements of people as well as maize; a number of South American maize races are clearly descended from populations introduced from the USA in the past 200 years.

A similar SSR survey of Old World and New World maize (Mir et al. 2013) revealed large-scale patterns of relationship and dissemination of maize between these regions. Their results indicate a large influence of Spanish and Portuguese trading and slave routes on the dispersion of maize from different regions of South America to Africa and Asia. European maize was descended from separate introductions of Northern Flint germplasm to northern Europe, various germplasm introductions of southeastern USA, Caribbean, and lowland tropical maize to southern Europe, and a region of admixture between Northern Flint and other types in the northern Iberian peninsula (Mir et al. 2013). These results are mostly congruent with other studies on the origin and diversity of maize in Europe, although the number and origin of independent introductions of tropical populations is debated (Revilla et al. 1998; Rebourg et al. 2001; Gauthier et al. 2002; Brandenburg et al. 2017).

20.1.4 Do Racial Classifications Based on Key Morphological Traits Accurately Reflect Phylogenetic Relationships?

Now that extensive data on molecular diversity and relationships among maize populations are becoming available, do they support previous classifications based on morphological traits? Sánchez et al. (2000a) directly compared the isozyme-based and the morphological/numerical taxonomy-based groupings of 209 accessions representing 59 Mexican races, finding a large-scale congruence of the two analyses. Three of four major racial complexes ('Central and Northern highlands,' 'Eight-rowed,' and 'Medium to low elevation' groups) identified with morphological data are also identified with isozyme data. The morphological data separated an additional small group of distinct specialty food types ('Chapalote' group) from the others, but these types were not distinct in the isozyme groupings. In general, groupings were less clear with isozyme data than morphological data, suggesting the morphological data exaggerates the genetic differences among groups. In addition, some other differences in groupings were identified between the two analyses, but the general picture is one of large-scale concordance with mostly finer-scale discrepancies between isozyme and morphological groupings. Subsequent analysis of the relationships among Mexican races with SSRs resulted in clear groups that mostly agreed with the morphological classifications, except for some races (Jala, Bolita, Harinoso de Ocho, and Maiz Dulce) with unique ear and kernel characteristics (Reif et al. 2006;

Bedoya et al. 2017). These races represent cases where strong human selection for specific characters used in racial classification coupled with genetic drift distorted the morphological relationships relative to the overall genetic relationships. Races have also been grouped on the basis of ecological variables from their geographic range, which also produces generally similar large-scale groupings (Ruiz Corral et al. 2008). Genetic, racial, and environmental variations are all distributed in a strongly correlated manner (Bretting et al. 1990; Sánchez et al. 2000b; Vigouroux et al. 2008; Arteaga et al. 2016).

20.1.5 Is Functional Genetic Variation Distributed in the Same Way as Neutral Genetic Variation?

Whereas F_{ST} measures genetic differentiation among subpopulations, Q_{ST} is an analogous measure of differentiation among subpopulations for quantitative traits (Spitze 1993). Theoretically, in the absence of selection or spatial patterning, Q_{ST} and F_{ST} have similar distributions, permitting tests of diversifying selection among subpopulations, although there are numerous caveats to such comparisons (Whitlock 2008). A few estimates of trait differentiation among subpopulation hierarchies in maize have been reported. Sánchez et al. (1993) did not report Q_{ST} directly, but did report variance components for 47 traits measured in 10 environments within and among 50 Mexican races represented by 162 accessions. Using these estimates, we computed the ratio of trait differentiation among races to differentiation among accessions within races (the value Q_{RT}/Q_{AR}), which has an average value of 0.85 across all trait groups (Table 20.4), indicating strong trait differentiation among Mexican races. This value can be compared to an isozyme-based genetic differentiation estimate for a similar set of populations made by Sánchez et al. (2000a), $F_{RT}/F_{AR} = 0.48$ (Table 20.4). Thus, differentiation among races

Table 20.4 Ratios of the variation (or differentiation) among races to variation (or differentiation) among accessions within races, and the ratio of relative trait differentiation among races to relative genetic differentiation among accessions within races. Trait variance components estimated by Sánchez et al. (1993) based on 162 accessions of 50 Mexican races measured for 47 traits measured in 10 environments. Isozyme-based genetic differentiation estimated for a similar set of populations by Sánchez et al. (2000a)

Character set	Ratio of variation among races to variation among accessions within races	Ratio of differentiation among to differentiation within races
		Q_{RT}/Q_{AR}
Vegetative	7.96	0.89
Agronomic	6.62	0.87
Tassel	6.35	0.86
Spikelet	3.02	0.75
Ear	4.9	0.83
Kernel	6.88	0.87
Cupule	3.74	0.79
Mean traits	5.87	0.85
		F_{RT}/F_{AR}
Isozymes		0.48
		$(Q_{RT}/Q_{AR})/(F_{RT}/F_{AR})$
Ratio of trait differentiation among races to genetic differentiation within races		1.78

Table 20.5 Genetic and phenotypic differentiation estimates for maize among villages within Mexican regions (Q_{VT} or F_{VT}), among populations within villages (Q_{PV} or F_{PV}), or overall individuals relative to total variation (Q_{IT} or F_{IT}), adapted from van Heerwaarden (2007) and Pressoir and Berthaud (2004a)

Altitude	Trait group	Among villages	Among populations within villages	Overall differentiation
		Q_{VT} or F_{VT}	Q_{PV} or F_{PV}	Q_{IT} or F_{IT}
Results from van Heerwaarden (2007)				
Highlands	SSRs	0.03	0.01	0.03
Highlands	Phenological traits	0.40	0.10	0.48
Highlands	Kernel traits	0.20	0.03	0.27
Highlands	All traits	0.30	0.06	0.35
Highlands	Average trait differentiation/Genetic differentiation	11.5	7.7	10.3
Lowlands	SSRs	0.03	0.06	0.09
Lowlands	Phenological traits	0.17	0.06	0.23
Lowlands	Kernel traits	0.13	0.22	0.33
Lowlands	All traits	0.15	0.14	0.28
Lowlands	Average trait differentiation/Genetic differentiation	5.4	2.2	3.2
Results from Pressoir and Berthaud (2004)				
Intermediate	SSRs	0.01	0.00	0.01
Intermediate	All traits	0.06	0.18	0.22
Intermediate	Average trait differentiation/Genetic differentiation	5.6	58.3	19.5

relative to within races is nearly twice as large for traits than random genetic markers (Table 20.4).

Another pair of studies compared differentiation among Mexican populations in different villages and among populations within villages (Pressoir and Berthaud 2004a; van Heerwaarden 2007) (Table 20.5). Both studies reported relatively little genetic differentiation overall; in the case of Pressoir and Berthaud (2004a), the low level of differentiation among populations was not surprising because sampling was only performed within a single race (Bolita). In the study by van Heerwaarden (2007), differentiation was estimated only within altitudinal ranges, so total variation among villages in high and low altitudes was not reported. In both cases, intermediate levels of differentiation were observed among and within villages, such that the ratio of differentiation among villages to differentiation within villages was much greater (by 2.2- to 58-fold) for traits than for genetic markers (Table 20.5).

Taken together, these results show a consistent pattern of stronger differentiation among races at the trait level than the level of random genetic markers. This implies 'functional' variation which is more strongly differentiated among races and populations than overall sequence variation. Human selection for local adaptation, as well as ear and seed type, has acted strongly on a subset of genetic variants in the evolution of maize races. If only a few genes were involved in the selected traits, this result would not be surprising. However, the largely polygenic nature of most maize

traits (Buckler et al. 2009; Schön et al. 2004; Laurie et al. 2004; Kump et al. 2011; Peiffer et al. 2014) suggests strong selective sweeps at a few loci cannot explain this observation. An alternative explanation is the low level of linkage disequilibrium expected in diverse open-pollinated maize populations (Remington et al. 2001) which allows selection to operate on relatively large sets of polygenic functional variants while leaving only diffuse genome-wide selection signals. Relatively small changes in allele frequency occurring at many loci result in surprisingly strong phenotypic differences.

Another possible explanation of the strong trait differentiation observed among maize races is that races were originally defined based on morphological traits, so by definition, they must be well-differentiated. This is certainly true for ear and kernel characters, and the geographic patterning of race distribution ensures that phenological traits should be differentiated among races. However, trait differentiation is observed quite consistently across groups of traits, some of which have little to do with racial classification (Table 20.3). As an extreme example, Holland and Goodman (1995) estimated variance among and within Latin American landrace accessions (mostly from different races) for grain yield combining ability on a common tester in temperate environments, finding twice as much variation observed among accessions than within accessions. Some of this differentiation may be due to genetic correlations with traits originally selected by farmers to create the local landrace populations, but we expect such correlations to be relatively weak. It seems more likely that relatively small allele frequency changes amplified across many loci can produce large effects on trait expression. An important implication of this finding for crop improvement is that sampling among accessions, and even more importantly, among races should be the highest priority for evaluating maize genetic resources for agronomically useful alleles. Sampling within accessions should be prioritized only after a wide sampling of races has been conducted, and the best sources of germplasm have been identified.

20.1.6 How Much Molecular Variation Exists Beyond SNPs: Transposable Elements, Repetitive DNA, Structural Variation, and Genome Size Variation?

About 85% of the maize genome is composed of repetitive sequences derived from transposable elements (Schnable et al. 2009; Jiao et al. 2017). An exhaustive comparison of the *bz* genomic region of eight maize haplotypes from different origins including two landraces (Coroico from the Amazon basin and another from Mexico) revealed remarkable variability in a ~100 kb region. The percentage of sequence shared between the two landraces was 40% while variability was found for many TE families (Wang and Dooner 2006).

Although we have only a limited understanding of the phenotypic effects of transposable element variation, there is good evidence that transposable elements have influenced evolution and domestication in crops (Oliver et al. 2013). A TE insertion in a regulatory region of the maize domestication gene *tb1* acts as an enhancer of gene expression and partially explains the increased apical dominance in maize (Studer et al. 2011). Similarly, a TE insertion in the promoter region of *gt1* (which controls the number of ears per plant) may be responsible for the differences in prolificacy between modern inbreds and some maize landrace populations (Wills et al. 2013).

ZmCCT9 and *10* are yet additional examples of the contribution of transposons to maize adaptation. Transposable element insertions in the regulatory regions of these genes suppress their expression, leading to earlier flowering under long daylengths (Hung et al. 2012; Yang et al. 2013). Both transposon insertions were found in high frequency in landraces grown at high latitude in the Americas, and these variants likely facilitated the spread of maize beyond its tropical zone of origin (Huang et al. 2017).

Comparison of TE composition in the genomes of inbred B73 (the standard maize reference genome) and the landraces Palomero Toluqueño and Olote Colorado (Diez et al. 2014) revealed some TEs are conserved across these distinct maize germplasm samples and likely represent core components of the *Zea* pan-genome. However, other individual TE subfamilies vary in copy number between landraces (Diez et al. 2014) and represent variable components of the pan-genome. Wang and Donner (2006) also demonstrated that TE composition, copy number, and physical arrangement within a single genomic region can vary significantly among unrelated inbred lines.

The size of the maize genome varies considerably among landraces, wild relatives, and inbred lines (Muñoz-Diez et al. 2012; Díez et al. 2013) with genome size negatively correlated with altitude and latitude (Díez et al. 2013; Laurie and Bennett 1985). Similar results were found in other studies of landraces and wild relatives, in which a negative correlation between genome size, heterochromatic chromosomal content, or specific repeat chromosomal knobs and altitude was often found, presumably due to a general relationship between genome size, DNA replication, and time to maturity (Realini et al. 2016; Fourastié et al. 2017; Bilinski et al. 2017; Jian et al. 2017).

20.2 Allele Mining in Maize Germplasm

Allele mining involves targeted re-sequencing of selected candidate loci in an attempt to uncover natural variation affecting a phenotype of interest. Allele mining can deliver a high proportion of functionally enriched data at relatively low cost and bioinformatic complexity, while increasing the ability to detect rare variants. A primary requirement for allele mining is knowledge of which genes are causally involved in important trait variation and therefore should be targeted for re-sequencing. Sequencing target genes in panels of diverse germplasm can reveal novel alleles with unknown functional effects. Association analyses of these alleles can provide initial estimates of the effects (and therefore the breeding utility) of these alleles. If alleles with effects superior to those present in elite breeding material are identified, they can be prioritized as targets for molecular breeding. Molecular markers based on or tightly linked to the putative causal polymorphisms of those alleles can then be designed and used to assist in the introgression of those alleles into elite cultivar genetic backgrounds. The effects of the alleles should be re-evaluated in elite genetic backgrounds to determine if their desired effects are consistent and no unfavorable pleiotropic or linked effects are observed. Once validated as useful, these new alleles and their associated allele-specific marker can be used in marker-assisted selection (MAS) to develop improved varieties. Allele mining can also be used as a form of pre-breeding to find a desired allele (or an equally useful variant) in more suitable breeding material than it was originally discovered. This facilitates introgression by reducing the chance of introducing unfavorable linked alleles. Allele mining has been more extensively employed in self-pollinating crops, particularly rice (Leung et al. 2015; Ashkani et al. 2015), although there are some examples of its use in maize, mostly for nutritional enhancement of maize grain (Harjes et al. 2008; Yan et al. 2010; Burt et al. 2011; Owens et al. 2014).

A technical hindrance to the use of allele mining is the limitation on acquiring extensive sequence data for targeted loci across diverse germplasm. Despite prodigious advances in sequencing technology, only two high-quality maize whole genome sequences have been published to date (Hirsch et al. 2016; Jiao et al. 2017) while at least eight other assemblies are nearly complete. This illustrates the continued expense and difficulty associated with whole genome sequencing (WGS) in the large, complex, and mostly repetitive genome of maize. Furthermore, this indicates WGS on multiple samples from population-based studies remains difficult, especially in highly heterogeneous and heterozygous landraces. While reduced representation sequencing strategies such as genotype-by-sequencing (GBS) excel at developing large numbers of

randomly distributed SNPs, they generally provide poor coverage of specific loci that may be of interest (Elshire et al. 2011; Glaubitz et al. 2014). Although imputation can be used to estimate missing sequence information, it is very unlikely to predict rare alleles (Hickey et al. 2012; Swarts et al. 2014).

20.2.1 Identifying Appropriate Targets for Allele Mining and Sources of Germplasm

Identifying loci for which allele mining is likely to be useful is occasionally straightforward, although these may be exceptional cases. For some traits, cloned genes known to affect the phenotype of interest in maize or other species are ideal starting points. From these known genes, the scope of re-sequencing targets can be widened to include paralogs, related genes, and genes known or expected to operate in the same pathway. Loci in close proximity to strong selective sweeps may be bountiful targets for mining as minor beneficial alleles could be in linkage repulsion with key domestication genes. Furthermore, alleles fixed directly at domestication loci by early maize breeders may no longer be ideal under modern production practices (Sood et al. 2014).

For many traits, however, no causally related genes are known. The converse problem can also occur: For some very complex and poorly understood traits, the list of genes or pathways potentially involved in their control can be hopelessly large. In these cases, some forward genetic analysis is required. QTL mapping in biparental families can be helpful, but unless the true genetic effects at any individual loci are very large, the QTL intervals are likely to be very large and the allele effects poorly estimated. Alternative linkage mapping approaches in maize include advanced intercross line populations, which have much higher mapping resolution than typical populations (Balint-Kurti et al. 2007), and multi-parental mapping populations, such as nested association mapping or multi-parental advanced generation intercross populations (McMullen et al. 2009; Buckler et al. 2009; Dell'Acqua et al. 2015; Holland 2015). If these high-resolution mapping resources are not segregating for the trait of interest, genome-wide association analysis can be conducted in diversity panels known to contain desirable trait variation (Brachi et al. 2011; Myles et al. 2009; Huang and Han 2014). Another excellent resource for gene discovery and allele mining are introgression libraries of near-isogenic lines (NILs) containing small genome segments introgressed from potentially multiple diverse parents into a common adapted genetic background (Tanksley and McCouch 1997; Zamir 2001; Fridman et al. 2004). There are a few publicly available NIL or advanced backcross introgression line sets in maize. Of particular interest for increasing the diversity of maize cultivars are the Germplasm Enhancement of Maize (see below) 'allelic diversity' lines containing landrace introgressions into elite temperate line backgrounds (Brenner et al. 2012; Sánchez et al. 2018) and a set of nearly 1000 lines containing introgressions from 10 teosinte accessions in the common B73 inbred genetic background (Lennon et al. 2016; Liu et al. 2016b; Karn et al. 2017). These two NIL sets have already uncovered novel genetic variation for disease resistance, kernel quality, water and nutrient uptake, and cell wall digestibility. The discovery of useful alleles in an adapted genetic background greatly facilitates its introgression into breeding programs.

If there are large effect variants segregating in the population, linkage and association studies may be able to resolve the causal gene, or at least define a few narrow intervals containing a handful of genes that can be targeted for allele mining (Harjes et al. 2008; Cook et al. 2012; Romay et al. 2013; Yang et al. 2017). For complex traits under mostly polygenic control, however, the likelihood that one or a few genes have large effects on trait variation is small. Unfortunately, in maize, most important agronomic traits are polygenic and most genome-wide association discoveries account for only a small proportion of trait variation (Schön et al.

2004; Laurie et al. 2004; Buckler et al. 2009; Kump et al. 2011; Zila et al. 2014). Worse, it is possible for highly significant associations to have overestimated effects (Larsson et al. 2013) or have limited repeatability across germplasm, environmental, or even marker samples (Bian et al. 2014). Highly polygenic traits are generally not good targets for allele mining.

Determining which germplasm to mine is another critical decision. While landraces and teosinte obviously serve as important sources of variation, choosing the correct subset to investigate is critical. There is likely little value in mining material that is closely related or contains alleles already prevalent in elite breeding material, but this information can be difficult to predict in advance. Allele mining likely offers the most potential for traits that have been mostly ignored by commercial breeders, have only recently become of interest, or are of evolutionary or anthropologic interest.

The reduction in genetic diversity that occurred during domestication may have eliminated some desirable traits; for example, maize seems to have lost some of teosinte's capacity to emit volatile signals that attract beneficial insects (Gouinguené et al. 2001; de Lange et al. 2016). A strategy to understand the consequence of domestication is by examining the diversity in genes of interest in both the domesticated and wild *Zea* gene pools. Such surveys have identified on the order of 1000 genes as showing strong signals of selection that reduce their diversity significantly below the average reduction that occurred genome-wide due to reduced effective population size during domestication (Wright et al. 2005; Hufford et al. 2012; Whitt et al. 2002). Inferences about why certain genes were targets of selection during domestication are often based on annotation information, but direct proof of their effects on phenotypes may require evaluation of the effects of wild alleles introgressed into domesticated genetic backgrounds (Wills et al. 2017).

Other traits that exist in landrace maize but have largely been lost from commercial inbred germplasm are food quality characters in which certain landraces are known to excel, such as the Orange Flints for carotenoid content (Burt et al. 2011), landraces from the highlands of Central Mexico for tortilla quality (Vázquez-Carrillo et al. 2011), landraces from the northwest of Spain for bakery products (Samayoa et al. 2016), and the grain anthocyanin and oil contents of landraces from the southwestern USA (Nankar et al. 2016). Another example is landraces native to areas with certain endemic biotic and abiotic stresses that can serve as useful repositories of tolerance and resistance alleles (Ruiz Corral et al. 2008). When allele mining from landraces or wild relatives, it may prove difficult to transfer novel alleles into suitable, adapted breeding material or to break linkages with other undesirable alleles. Therefore, extensive pre-breeding may be required after identification of target alleles. The advent and implementation of genome editing tools, specifically the CRISPR/CRISPR-associated protein 9 (Cas9) system, has potential to accelerate the delivery of exotic-derived alleles or possibly even specific gain-of-function gene additions into elite materials while avoiding backcrossing generations and linkage drag (Liu et al. 2016a).

20.2.2 Allele Mining Technologies

Traditional Sanger sequencing is a suitable, low-cost approach for limited allele mining (≤ 5 genes) in a small number of genotypes (≤ 50). While primer design, amplification, and sequencing can be easily completed for coding regions using relatively generic laboratory equipment and software, ThermoFisher Scientific offers a 'Targeted Sequencing by Sanger Sequencing' product combination to facilitate the process. Sequence amplification outside of coding regions is substantially more difficult because of the higher levels of diversity, copy number variation, and re-arrangement. Sanger sequencing can identify SNPs and small InDels with high accuracy but lacks the scalability of the approaches described below. Nevertheless, allele mining via Sanger has been performed for carotenoid and stress response genes in maize (Burt et al. 2011; Estermann et al.

2017). This type of limited re-sequencing is also commonly part of traditional gene cloning studies where the causal gene is subsequently sequenced in a diverse panel of material (i.e., Yang et al., 2013). However, the intent here is generally to show association between genotype and phenotype, rather than to detect novel variation.

The earliest high-throughput methods for allele mining were based on hybridization and microarrays. In this approach, complementary probes were designed that tiled across the region of interest and then fixed to an array. Fragmented genomic DNA was then washed over the array, and complementary sequences were pulled down by hybridization. Following release of bound fragments, sequencing was performed using second generation machines like the Roche 454. The most notable of the microarrays was the since discontinued Roche NimbleGen Sequence Capture Microarray. This product was successfully used to mine 4,648 biomass production and composition genes in 21 diverse maize inbreds (Muraya et al. 2015). NimbleGen was also used to re-sequence 43 genes and a 2.2 Mb region in B73 and Mo17 following a proof-of-concept of a novel method for removal of repetitive DNA (Fu et al. 2010).

Following the decline of the microarray, hybridization-based approaches shifted to a solution- or liquid-based capture method in which probes float freely rather than being fixed to an array. In theory, this should increase the probability of hybridization. Probes carry a label (i.e., biotin) that allows captured fragments to be easily collected (i.e., via avidin or streptavidin) prior to library prep. Sequencing occurs on current generation systems such as Pacific Biosciences SMRT or Oxford Nanopore. Commercially available versions of solution- based hybridization that are readily amenable to maize include Roche NimbleGen SeqCap EZ, Agilent SureSelect, and IDT xGen Lockdown.

While hybridization-based approaches excel at investigating large numbers of genes (50+) simultaneously, they suffer from a number of drawbacks such as requiring a large amount (1–3 μg) of high-quality DNA per sample, issues with non-specificity or off-target hybridization (especially due to paralogs and secondary capture), and low sequencing coverage. Therefore, multiplex PCR-based methods (also known as amplicon-based enrichment) have been developed that attempt to alleviate these concerns. In this approach, barcoded primers are designed to cover a region of interest. Multiplexed primers are then used to amplify the selected regions prior to pooling, library prep, and current generation sequencing. Examples for use in maize include ThermoFisher Ampliseq, Illumina TruSeq, and Roche HEAT-Seq. Ampliseq was used to design 319 PCR assays covering ~86 kb of 20 key photoperiod response loci in 95 diverse maize samples followed by ion torrent semiconductor sequencing (Jamann et al. 2017). However, a major impediment to PCR-based approaches is the inherent limits of multiplexing due to primer interaction and the need to optimize all primer pairs in order to avoid PCR bias.

Numerous issues plague both hybridization- and PCR-based approaches to allele mining as well. Chief among them is the need to balance stringency versus loci coverage when designing probes or primers. Neither approach is capable of covering the majority of loci with a single primer or probe. Therefore, multiple probes or primers are needed per loci. Invariably, repetitive or high homology regions will be encountered that are difficult to design specific probes or primers for. The choice must then be made whether it is better to omit mining the repetitive region of the locus or risk acquiring off-target sequencing reads. Both approaches also suffer from ascertainment bias as probes or primers are designed based on reference sequences and may fail to bind in diverse germplasm. Neither approach is particularly good at detecting large structural variation or InDels greater than a few hundred base pairs. It should also be noted that most current allele mining techniques were developed for use in human genetics and have only rarely been validated in maize. Furthermore, the field remains highly fluid with many technologies and strategies likely to rapidly appear (and disappear) over the coming years until the market coalesces around an ideal platform(s).

20.3 Using Exotic Germplasm to Enhance Polygenic Traits

Commercial breeding programs focus on short-term gains for yield and have limited incorporation of tropical and exotic materials into their breeding pipeline. This is mainly due to an increase in the number of breeding cycles and the associated linkage drag with traditional means of introgression for the inclusion of these materials (Crossa and Gardner 1987; Bernardo 2009). Furthermore, these types of programs tend to recycle lines for genetic gain instead of maintaining higher levels of variation from lines developed from improved populations or synthetics (Pollak 2001). Short-term-oriented breeding programs have continued to produce high yields, but yield plateaus can be experienced due to erosion of variability (Grassini et al. 2013). USA's maize breeding and production is primarily dependent on pedigrees that trace to relatively few adapted lines such as B73, A632, B14, Oh43, C103, and Mo17—a small sample of the allelic diversity available to USA breeders (Pollak 2001; Goodman 2005; Mikel and Dudley 2006; Smith 2007; Nelson et al. 2008). These low levels of diversity for a widely grown monoculture crop create vulnerability to shifting pathogen populations (Ullstrup 1972; Michelini and Hallauer 1993) and limit yield gains.

As described earlier, most important agronomic traits in maize are controlled by many genes, mostly with very small effects. The polygenic nature of most traits severely limits the effectiveness of allele mining, gene editing, and introgression breeding approaches to improving quantitative traits with exotic germplasm. Traditional breeding methods such as recurrent selection and pedigree breeding with inbred and hybrid testing are efficient at improving polygenic traits because they select on the combined value of all the alleles carried in breeding lines. These methods have been applied to breeding with exotic maize germplasm in temperate regions with some success (Hallauer and Sears 1972; Uhr and Goodman 1995; Tallury and Goodman 1999; Goodman 2005; Nelson and Goodman 2008; Ron Parra and Hallauer 2010; Hallauer and Carena 2014). One relatively small scale but long-term public breeding program directed by Dr. Major Goodman at North Carolina State University has developed inbreds from purely tropical origins representing an entirely new heterotic group with yield potential comparable to modern commercial hybrids, at least in the southeastern USA (Tallury and Goodman 1999; Goodman 2005). Such results demonstrate the potential utility of exotic germplasm for improving commercial temperate cultivars.

The long-term potential utility for tropical maize germplasm to improve disease resistance and yield is offset by their tendency for maladaptation due to photoperiod sensitivity, weak stalks, poorly developed root structures, and relatively low agronomic performance per se (Hallauer and Sears 1972; Albrecht and Dudley 1987; Holland and Goodman 1995). Furthermore, landrace collections that have not been subjected to selection under inbreeding have very high levels of inbreeding depression that limit their usefulness for breeding. Highland maize races, in particular, have very specific adaptation that severely restricts their potential to improve yields in temperate (or even tropical lowland) environments (Goodman 2005). These are among the least promising sources of tropical germplasm for breeding polygenic traits. Much better sources of germplasm are tropical inbred lines, hybrids, or pre-adapted populations (Goodman 2005).

The Latin American Maize Project (LAMP) was the first coordinated international project for evaluating native germplasm collections from over 12 countries (Pollak 2001), and the results from this project allowed selection of the most promising tropical populations, based on performance in their zone of adaptation. Building on this program, the Germplasm Enhancement of Maize Project was established to incorporate the best exotic maize germplasm into the temperate USA breeding pool. The project is a collaborative effort of the USDA, university partners, and private industry breeding programs (Goodman

2005). With industry cooperation, tropical germplasm sources are crossed to proprietary inbreds, and in some cases backcrossed to another proprietary inbred of the same heterotic group, followed by selection for inbred performance per se and topcross performance on modern Corn Belt Dent testers. The project has released 295 early generation-derived inbred lines to date (http://www.public.iastate.edu/~usdagem/GEM_Project/GEM_Project.htm), representing a major influx of new alleles into the temperate breeding pool.

20.3.1 Pre-breeding Strategies for Exotic Resource Incorporation and Introgression

Pre-breeding includes the introduction, adaptation, evaluation, and improvement of germplasm resources for use in a breeding program, mainly the use of tropical and/or exotic lines for the increase of genetic variability and additional heterotic vigor, with the goal of lessening the chances for a yield plateau. The genetic base of most maize breeding programs can be enhanced by new alleles from exotic parents either by introgression or incorporation (Simmonds 1993). Introgression is described by the process of backcrossing alleles from exotic parents into elite lines. The alleles typically have large effects on qualitative traits (e.g., disease resistance genes) and continuous cycles of backcrossing and selection can preserve the elite parental genome while introgressing exotic alleles. In contrast to introgression, incorporation is a method by which populations are developed locally to slowly adapt exotic materials for use as parental lines in a crossing program (Hallauer and Sears 1972; Holland 2004). These populations are under selection for gradual improvement over numerous breeding cycles. Incorporation strategies are more likely to be useful in the long-term for improving polygenic traits.

20.3.2 Applying Genomic Selection to Exotic Germplasm

Genomic selection (GS) is a method to incorporate information from genome-wide markers to predict the breeding values of individuals regardless of whether they have been phenotyped or not (Meuwissen et al. 2001). Rather than trying to identify the most important markers associated with traits, as is the goal of quantitative trait locus or association mapping, GS assumes that all (or at least very many) markers are associated with the trait, as expected for highly polygenic traits (Bernardo and Yu 2007; Heffner et al. 2009). This assumption, along with statistical models that allow the use of many more markers than phenotypic observations for making predictions, allows GS to achieve reasonably good prediction accuracy. Good prediction accuracy, coupled with the capacity to conduct selections on individual plants or seeds and in additional generations per year, allows GS to achieve higher gains per unit time than direct phenotypic selection under the right circumstances (Heffner et al. 2010).

GS could become a powerful tool in the improvement of polygenic traits using exotic germplasm. Increases in the speed and efficiency of introgression can be attained through genomic selection (Jacobson et al. 2015; Bernardo 2016). Typically, gains from selection in an adapted by exotic cross are made gradually and can take between 5 and 10 cycles to achieve a significant level of success. These cycles can be drastically reduced (two- to threefold) by implementing GS. Achieving such significant increases in improvement per unit time will require the use of full-season nurseries or greenhouses to increase the number of generations per year (Bernardo 2009).

Landrace populations can be selected through GS to not only improve the efficiency of local adaptation among unadapted populations but also identify allelic contributions for specific improvements to adapted lines (Dwivedi et al. 2016; Gorjanc et al. 2016). Genomic selection

models applied to diverse germplasm collections of sorghum have shown wide generality (Yu et al. 2016); however, this may be less likely in maize because of the higher diversity and lower linkage disequilibrium. Initial studies on the predictive ability of genomic selection models across populations in maize have indicated limited transferability of predictive accuracy across diverse germplasm sets (Windhausen et al. 2012; Peiffer et al. 2014).

Although Bernardo (2009) suggested that GS can significantly increase gain from selection per unit time in exotic by adapted cross populations, it is likely that much of this gain would come from simply reselecting the adapted alleles. Other simulations suggest that such a result is likely, and therefore, it is likely better to implement genomic selection in pure exotic populations first (Gorjanc et al. 2016). We propose here a plan to first implement rapid genomic selection purely for adaptation as the initial phase in a long-term genomic-selection-enabled program (Fig. 20.2). Adaptation through selection for earlier flowering is relatively easy to accomplish based on single-plant phenotypic selection, although it can be slow because selection can only be practiced in the target environment (Hallauer and Sears 1972; Hallauer and Carena 2014; Teixeira et al. 2015). GS could help speed this process by first training the prediction models on data from the initial population grown in the target environment and evaluated primarily for flowering time, then executing GS multiple generations per year on a single-plant basis (Fig. 20.2).

Following adaptation of the pure exotic population, individuals or lines can be crossed to an elite temperate tester to form topcross hybrids for multi-environment yield evaluations (Fig. 20.2). In general, tropical maize combines about equally well with any temperate heterotic group, although there are occasional indications of specific combining ability being important in these crosses (Crossa et al. 1987; Beck et al. 1991; Holland and Goodman 1995). We have indicated two predominant commercial temperate heterotic groups (e.g., Stiff Stalk and non-Stiff Stalk), but the method outlined in Fig. 20.2 could be implemented with more temperate groups. The yield data can be joined with marker information on the exotic parents of the hybrids to generate a new prediction model for topcross agronomic performance, and could involve an index of multiple important traits measured in the yield trials (such as yield, grain dry-down, and lodging). This prediction model could then be applied in multiple generations per year on a single-plant basis to rapidly improve the combining ability of the tropical population with the desired tester.

The final phase of population improvement could be the development of semi-exotic populations by crossing the adapted and improved pure tropical population to lines from a different temperate heterotic group than was used for topcross evaluation (Fig. 20.2). This new population could then be tested again for combining ability to develop a third prediction model that could be implemented to develop a new improved 50% exotic population, from which inbreds could be extracted for testing and selection as hybrid parents. In all three of these genomic selection phases, we have not specified the number of cycles, but breeders should be aware that prediction accuracy of the models breaks down over cycles of recombination. Therefore, after three or four cycles, if the desired level of performance has not been achieved, the models probably should be retrained on the latest generation before continuing GS (Müller et al. 2017).

The proposed breeding scheme outlined in Fig. 20.2 is untested and should be considered a proposal only. Surely other approaches can be taken to incorporate exotic germplasm into elite breeding pools. Breeders should consider that phenotypic selection within pure exotic populations for adaptation and combining ability with adapted tester lines has succeeded to a greater extent than molecular-aided backcrossing approaches. Therefore, we recommend the use of molecular data as an accompaniment to population-focused breeding, rather than single-gene-focused breeding. Finally, breeding success is more likely within improved exotic materials, but perhaps the promise of increased rates of genetic gains from genomic selection will

help make landrace populations a more useful source of favorable alleles for maize breeders.

References

Aguilar J, Illsley C, Marielle C (2003) El maize como cultivo I. In: Esteva G, Marielle C (eds) Sin Maiz No Hay Pais. Consejo Nacional para la Cultura y las Artes, pp 83–122

Albrecht B, Dudley JW (1987) Evaluation of four maize populations containing different proportions of exotic germplasm. Crop Sci 27:480. https://doi.org/10.2135/cropsci1987.0011183X002700030011x

Anderson E (1946) Maize in Mexico a preliminary survey. Ann Mo Bot Gard 33:147–247. https://doi.org/10.2307/2394428

Anderson E, Cutler HC (1942) Races of Zea mays: I. Their recognition and classification. Ann Mo Bot Gard 29:69–88. https://doi.org/10.2307/2394331

Arteaga MC, Moreno-Letelier A, Mastretta-Yanes A et al (2016) Genomic variation in recently collected maize landraces from Mexico. Genomics Data 7:38–45. https://doi.org/10.1016/j.gdata.2015.11.002

Ashkani S, Rafii Yusop M, Shabanimofrad M, et al (2015) Allele mining for blast resistance in rice. Curr Issues Mol Biol 17:57–74. https://doi.org/10.21775/cimb.017.057

Azar C, Mather DE, Hamilton RI (1997) Maize landraces of the St. Lawrence—Great Lakes region of North America. Euphytica 141–148. https://doi.org/10.1023/a:1003035231969

Balint-Kurti PJ, Zwonitzer JC, Wisser RJ et al (2007) Precise mapping of quantitative trait loci for resistance to southern leaf blight, caused by *Cochliobolus heterostrophus* race O, and flowering time using advanced intercross maize lines. Genetics 176:645–657. https://doi.org/10.1534/genetics.106.067892

Beck DL, Vasal SK, Crossa J (1991) Heterosis and combining ability among subtropical and temperate intermediate-maturity maize germplasm. Crop Sci 31:68–73. https://doi.org/10.2135/cropsci1991.0011183X002600010017x

Bedoya CA, Dreisigacker S, Hearne S et al (2017) Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. PLoS One 12:e0173488. https://doi.org/10.1371/journal.pone.0173488

Bellon MR (1991) The ethnoecology of maize variety management: a case study from Mexico. Hum Ecol 19:389–418. https://doi.org/10.1007/BF00888984

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49:419–425. https://doi.org/10.2135/cropsci2008.08.0452

Bernardo R (2016) Genomewide predictions for backcrossing a quantitative trait from an exotic to an adapted line. Crop Sci 56:1067–1075. https://doi.org/10.2135/cropsci2015.09.058

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082. https://doi.org/10.2135/cropsci2006.11.0690

Bian Y, Yang Q, Balint-Kurti PJ et al (2014) Limits on the reproducibility of marker associations with southern leaf blight resistance in the maize nested association mapping population. BMC Genomics 15:1068. https://doi.org/10.1186/1471-2164-15-1068

Bilinski P, Albert PS, Berg JJ et al (2017) Parallel altitudinal clines reveal adaptive evolution of genome size in *Zea mays*. PLOS Genet14:e1007162. https://doi.org/10.1371/journal.pgen.1007162

Bird RM, Goodman MM (1977) The races of maize. V. Grouping maize races on the basis of ear morphology. Econ Bot 31:471–481

Bracco M, Lia VV, Hernández JC et al (2012) Genetic diversity of maize landraces from lowland and highland agro-ecosystems of Southern South America: implications for the conservation of native resources. Ann Appl Biol 160:308–321. https://doi.org/10.1111/j.1744-7348.2012.00544.x

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol 12:232. https://doi.org/10.1186/gb-2011-12-10-232

Brandenburg J-T, Mary-Huard T, Rigaill G et al (2017) Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLOS Genet 13:e1006666. https://doi.org/10.1371/journal.pgen.1006666

Brenner EA, Blanco M, Gardner C, Lübberstedt T (2012) Genotypic and phenotypic characterization of isogenic doubled haploid exotic introgression lines in maize. Mol Breed 30:1001–1016. https://doi.org/10.1007/s11032-011-9684-5

Bretting P, Goodman MM, Stuber CW (1990) Isozymatic variation in Guatemalan races of maize. Am J Bot 77:211–225

Bretting PK, Goodman MM, Stuber CW (1987) Karyological and isozyme variation in West Indian and allied American mainland races of maize. Am J Bot 74:1601–1613. https://doi.org/10.1002/j.1537-2197.1987.tb08760.x

Brush SB, Perales HR (2007) A maize landscape: ethnicity and agro-biodiversity in Chiapas Mexico. Agric Ecosyst Environ 121:211–221. https://doi.org/10.1016/J.AGEE.2006.12.018

Buckler ES, Holland JB, Bradbury PJ et al (2009) The genetic architecture of maize flowering time. Science 325:714–718. https://doi.org/10.1126/science.1174276

Burt AJ, Grainger CM, Smid MP et al (2011) Allele mining of exotic maize germplasm to enhance macular carotenoids. Crop Sci 51:991–1004. https://doi.org/10.2135/cropsci2010.06.0335

Chávez E (1913) Cultivo del maiz. Secretaria de Fomento. Direccion General de Agricultura. Mexico, D.F.

Cook JP, McMullen MD, Holland JB et al (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158:824–834. https://doi.org/10.1104/pp.111.185033

Crawford GW, Saunders D, Smith DG (2006) Pre-contact maize from Ontario, Canada: context, chronology, variation, and plant association. In: Staller J, Tykot RH, Benz B (Eds). Histories of maize: multidisciplinary approaches to the prehistory, linguistics, biogeography, domestication, and evolution of maize. Academic Press, San Diego, pp 549–559

Crosby AW (1972) The Columbian exchange: biological and cultural consequences of 1492. Greenwood Press, Westport, CT

Crossa J, Gardner C (1987) Introgression of an exotic germplasm for improving an adapted maize population. Crop Sci 27:187–190. https://doi.org/10.2135/cropsci1987.0011183X002700020008x

Crossa J, Gardner CO, Mumm RF (1987) Heterosis among populations of maize (*Zea mays* L.) with different levels of exotic germplasm. Theor Appl Genet 73:445–450. https://doi.org/10.1007/BF00262514

de Lange ES, Farnier K, Gaudillat B, Turlings TCJ (2016) Comparing the attraction of two parasitoids to herbivore-induced volatiles of maize and its wild ancestors, the teosintes. Chemoecology 26:33–44. https://doi.org/10.1007/s00049-015-0205-6

Dell'Acqua M, Gatti DM, Pea G et al (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome Biol 16:167. https://doi.org/10.1186/s13059-015-0716-z

Díez CM, Gaut BS, Meca E et al (2013) Genome size variation in wild and cultivated maize along altitudinal gradients. New Phytol 199:264–276. https://doi.org/10.1111/nph.12247

Diez CM, Meca E, Tenaillon MI, Gaut BS (2014) Three groups of transposable elements with contrasting copy number dynamics and host responses in the maize (*Zea mays* ssp. *mays*) genome. PLoS Genet 10:e1004298. https://doi.org/10.1371/journal.pgen.1004298

Doebley J, Goodman MM, Stuber CW (1984) Isoenzymatic variation in *Zea* (*Gramineae*). Syst Bot 9:203–218. https://doi.org/10.2307/2418824

Doebley JF, Goodman MM, Stuber CW (1985) Isozyme variation in the races of maize from Mexico. Am J Bot 72:629–639. https://doi.org/10.2307/2443674

Doebley JF, Goodman MM, Stuber CW (1986) Exceptional genetic divergence of northern flint corn. Am J Bot 73:64. https://doi.org/10.2307/2444278

Doebley JF, Goodman MM, Stuber CW (1983) Isozyme variation in maize from the southwestern United States: taxonomic and anthropological implications. Maydica 28:97–120

Dubreuil P, Warburton M, Chastanet M et al (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291

Dwivedi SL, Ceccarelli S, Blair MW et al (2016) Landrace germplasm for improving yield and abiotic stress adaptation. Trends Plant Sci 21:31–42

Dyer GA, López-Feldman A, Yúnez-Naude A, Taylor JE (2014) Genetic erosion in maize's center of origin. Proc Natl Acad Sci USA 111:14094–14099. https://doi.org/10.1073/pnas.1407033111

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Estermann MA, Ugarte A, Pretini N et al (2017) Allele mining in the Argentine public maize inbred line collection of two paralogous genes encoding NAC domains. Mol Breed 37. https://doi.org/10.1007/s11032-016-0602-8

Fourastié MF, Gottlieb AM, Poggio L, González GE (2017) Are cytological parameters of maize landraces (*Zea mays* ssp. *mays*) adapted along an altitudinal cline? J Plant Res 1–12. https://doi.org/10.1007/s10265-017-0996-3

Fridman E, Carrari F, Liu Y-S et al (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305:1786–1789. https://doi.org/10.1126/science.1101666

Fu Y, Springer NM, Gerhardt DJ et al (2010) Repeat subtraction-mediated sequence capture from a complex genome. Plant J 62:898–909. https://doi.org/10.1111/j.1365-313X.2010.04196.x

Gauthier P, Gouesnard B, Dallard J et al (2002) RFLP diversity and relationships among traditional European maize populations. Theor Appl Genet 105:91–99. https://doi.org/10.1007/s00122-002-0903-7

Girola CD (1919) Maices argentinos y aclimatados. Talleres Graficos J. Weiss y Preusche, Buenos Aires

Glaubitz JC, Casstevens TM, Lu F et al (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346. https://doi.org/10.1371/journal.pone.0090346

Goodman MM (2005) Broadening the genetic base of US maize with exotic germplasm. Maydica 50:203–214

Goodman MM, Bird RM (1977) The races of maize. IV. Tentative grouping of 219 Latin American races. Econ Bot 31:204–221

Goodman MM, Brown WL (1988) Races of corn. In: Sprage GF, Dudley JW (eds) Corn and corn improvement, 3rd edn. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, pp 33–79

Goodman MM, Stuber CW (1983a) Races of maize. VI. Isozyme variation among races of maize in Bolivia. Maydica 28:169–187

Goodman MM, Stuber CW (1983b) Maize. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding. Elsevier, pp 1–33

Gorjanc G, Jenko J, Hearne SJ, Hickey JM (2016) Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace

populations. BMC Genomics 17:30. https://doi.org/10.1186/s12864-015-2345-z

Gouesnard B, Dallard J, Panouillé A, Boyat A (1997) Classification of French maize populations based on morphological traits. Agronomie 17:491–498. https://doi.org/10.1051/agro:19970906

Gouinguené S, Degen T, Turlings TCJ (2001) Variability in herbivore-induced odour emissions among maize cultivars and their wild ancestors (teosinte). Chemoecology 11:9–16. https://doi.org/10.1007/PL00001832

Grassini P, Eskridge KM, Cassman KG (2013) Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat Commun 4:2918. https://doi.org/10.1038/ncomms3918

Grobman A, Salhuana W, Sevilla R, Manglesdorf PC (1961) Races of Maize in Peru: Their Origins, Evolution and Classification. National Academy of Sciences—National Research Council, Washington, D. C

Hallauer A, Sears J (1972) Integrating exotic germplasm into Corn Belt maize breeding programs. Crop Sci 12:203–206. https://doi.org/10.2135/cropsci1972.0011183X001200020016x

Hallauer AR, Carena MJ (2014) Adaptation of tropical maize germplasm to temperate environments. Euphytica 196:1–11. https://doi.org/10.1007/s10681-013-1017-9

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One 2:e1367. https://doi.org/10.1371/journal.pone.0001367

Hamrick JL, Godt MJW (1997) Allozyme diversity in cultivated crops. Crop Sci 37:26–30. https://doi.org/10.2135/cropsci1997.0011183X003700010004x

Harjes CE, Rocheford TR, Bai L et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333. https://doi.org/10.1126/science.1150255

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50:1681. https://doi.org/10.2135/cropsci2009.11.0662

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1. https://doi.org/10.2135/cropsci2008.08.0512

Hernández Xolocotzi E (1972) Consumo humano de maiz y el aprovechamiento de tipos con alto valor nutritivo. In: Memoria del simposio sobre desarrollo y utilizacion de maices de alto valor nutritivo. Colegio de Postgraduados, Chapingo, Mexico, pp 149–156

Hernández Xolocotzi E (1985) Maize and man in the greater southwest. Econ Bot 39:416–430. https://doi.org/10.1007/BF02858749

Hickey JM, Crossa J, Babu R, de los Campos G (2012) Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. Crop Sci 52:654. https://doi.org/10.2135/cropsci2011.07.0358

Hirsch CN, Hirsch CD, Brohammer AB et al (2016) Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. Plant Cell 28:2700–2714. https://doi.org/10.1105/tpc.16.00353

Holland JB (2015) MAGIC maize: a new resource for plant genetics. Genome Biol 16:163. https://doi.org/10.1186/s13059-015-0713-2

Holland JB (2004) Breeding: incorporation of exotic germplasm. In: Goodman, RM (ed) Encyclopedia of Plant and Crop Science. Marcel Dekker, NY, p. 222–224

Holland JB, Goodman MM (1995) Combining ability of tropical maize accessions with U.S. germplasm. Crop Sci 35:767–773. https://doi.org/10.2135/cropsci1995.0011183X003500030022x

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nat Rev Genet 10:639–650. https://doi.org/10.1038/nrg2611

Huang C, Sun H, Xu D et al (2017) ZmCCT9 enhances maize adaptation to higher latitudes. Proc Natl Acad Sci USA 201718058. https://doi.org/10.1073/pnas.1718058115

Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol 65:531–551. https://doi.org/10.1146/annurev-arplant-050213-035715

Hufford, M (2010) Genetic and ecological approaches to guide conservation of teosinte (Zea mays ssp. parviglumis), the wild progenitor of maize. University of California, Davis

Hufford MB, Xu X, van Heerwaarden J et al (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811. https://doi.org/10.1038/ng.2309

Hugo PR, Brush SB, Qualset CO (2009) Dynamic management of naize landraces in central Mexico. https://doi.org/10.1663/0013-0001(2003)057[0021:DMOMLI]2.0.CO;2

Hung H-Y, Shannon LM, Tian F et al (2012) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc Natl Acad Sci USA 109:E1913–E1921. https://doi.org/10.1073/pnas.1203189109

Iltis HH, Doebley JF (1980) Taxonomy of *Zea* (*Gramineae*). II. Subspecific categories in the Zea mays complex and a generic synopsis. Source Am J Bot 67:994–1004

Jacobson A, Lian L, Zhong S, Bernardo R (2015) Minimal loss of genetic diversity after genomewide selection within biparental maize populations. Crop Sci 55:783–789. https://doi.org/10.2135/cropsci2014.10.0744

Jamann TM, Sood S, Wisser RJ, Holland JB (2017) High-throughput resequencing of maize landraces at genomic regions associated with flowering time. PLoS One 12:1–16. https://doi.org/10.1371/journal.pone.0168910

Jian Y, Xu C, Guo Z et al (2017) Maize (*Zea mays* L.) genome size indicated by 180-bp knob abundance is associated with flowering time. Sci Rep 7:1–9. https://doi.org/10.1038/s41598-017-06153-8

Jiao Y, Peluso P, Shi J et al (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527. https://doi.org/10.1038/nature22971

Jones ES, Sullivan H, Bhattramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115:361–371. https://doi.org/10.1007/s00122-007-0570-9

Karn A, Gillman JD, Flint-Garcia SA (2017) Genetic analysis of teosinte alleles for kernel composition traits in maize. G3 Genes Genomes Genetics 7:1157–1164. https://doi.org/10.1534/g3.117.039529

Kuleshov NN (1930) The maize of Mexico, Guatemala, and Colombia. Bull Appl Bot Genet Plant Breed 47:40–51

Kuleshov NN (1933) World diversity of phenotypes of maize. J Am Soc Agron 688–700

Kumar A, Kumari J, Rana JC et al (2015) Diversity among maize landraces in North West Himalayan region of India assessed by agro-morphological and quality traits. Indian J Genet Plant Breed 75:188–195. https://doi.org/10.5958/0975-6906.2015.00029.2

Kump KL, Bradbury PJ, Wisser RJ et al (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–168. https://doi.org/10.1038/ng.747

Larsson SJ, Lipka AE, Buckler ES (2013) Lessons from Dwarf8 on the strengths and weaknesses of structured association mapping. PLoS Genet 9:e1003246. https://doi.org/10.1371/journal.pgen.1003246

Laurie CC, Chasalow SD, LeDeaux JR et al (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141–2155. https://doi.org/10.1534/genetics.104.029686

Laurie DA, Bennett MD (1985) Nuclear DNA content in the genera *Zea* and *Sorghum*. Intergeneric, interspecific and intraspecific variation. Heredity (Edinb) 55:307–313. https://doi.org/10.1038/hdy.1985.112

Lennon JR, Krakowsky M, Goodman M et al (2016) Identification of alleles conferring resistance to gray leaf spot in maize derived from its wild progenitor species teosinte. Crop Sci 56:209–218. https://doi.org/10.2135/cropsci2014.07.0468

Leung H, Raghavan C, Zhou B et al (2015) Allele mining and enhanced genetic recombination for rice breeding. Rice 8:34. https://doi.org/10.1186/s12284-015-0069-y

Liu D, Hu R, Palla KJ et al (2016a) Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. Curr Opin Plant Biol 30:70–77

Liu K, Goodman M, Muse S et al (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Liu Z, Cook J, Melia-Hancock S et al (2016b) Expanding maize genetic resources with predomestication alleles: maize–teosinte introgression populations. Plant Genome 9:1–11. https://doi.org/10.3835/plantgenome 2015.07.0053

Logan AL, Hastorf CA, Pearsall DM (2012) "Let's drink together": early ceremonial use of maize in the Titicaca Basin. Lat Am Antiq 23:235–258. https://doi.org/10.7183/1045-6635.23.3.235

Lu H, Bernardo R (2001) Molecular marker diversity among current and historical maize inbreds. Theor Appl Genet 103:613–617. https://doi.org/10.1007/PL00002917

Matsuoka Y, Mitchell SE, Kresovich S et al (2002a) Microsatellites in *Zea*-variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450

Matsuoka Y, Vigouroux Y, Goodman MM et al (2002b) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084. https://doi.org/10.1073/pnas.052125199

Mayr E (1982) The growth of biological thought. Belknap Press of Harvard University, Cambridge, MA

McMullen MD, Kresovich S, Villeda HS et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740. https://doi.org/10.1126/science.1174320

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Michelini LA, Hallauer AR (1993) Evaluation of exotic and adapted maize (*Zea mays* L.) germplasm crosses. Maydica

Mikel MA, Dudley JW (2006) Evolution of North American dent corn from public to proprietary germplasm. Crop Sci 46:1193. https://doi.org/10.2135/cropsci2005.10-0371

Mir C, Zerjal T, Combes V et al (2013) Out of America: tracing the genetic footprints of the global diffusion of maize. Theor Appl Genet 126:2671–2682. https://doi.org/10.1007/s00122-013-2164-z

Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochem Cycles 22. https://doi.org/10.1029/2007gb002947

Müller D, Schopp P, Melchinger AE (2017) Persistency of prediction accuracy and genetic gain in synthetic populations under recurrent genomic selection. G3 Genes Genomes Genetics 7:801–811. https://doi.org/10.1534/g3.116.036582

Muñoz-Diez C, Vitte C, Ross-Ibarra J et al (2012) Using nextgen sequencing to investigate genome size variation and transposable element content. In: Grandbastien M-A, Casacuberta JM (eds) Plant transposable elements. Springer, Berlin, Heidelberg, pp 164–171

Muraya MM, Schmutzer T, Ulpinnis C et al (2015) Targeted sequencing reveals large-scale sequence polymorphism in maize candidate genes for biomass production and composition. PLoS One 10:e0132120. https://doi.org/10.1371/journal.pone.0132120

Myles S, Peiffer J, Brown PJ et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202. https://doi.org/10.1105/tpc.109.068437

Nankar A, Grant L, Scott P, Pratt RC (2016) Agronomic and kernel compositional traits of blue maize landraces from the southwestern United States. Crop Sci 56:2663–2674. https://doi.org/10.2135/cropsci2015.12.0773

Ndiso JB, Mugo S, Kibe AM, Pathaka RS (2013) Phenotypic diversity in local coastal maize landraces in Kenya. Int J Agric Sci 3:53–59

Nelson PT, Coles ND, Holland JB et al (2008) Molecular characterization of maize inbreds with expired US plant variety protection. Crop Sci 48:1673–1685. https://doi.org/10.2135/cropsci2008.02.0092

Nelson PT, Goodman MM (2008) Evaluation of elite exotic maize inbreds for use in temperate breeding. Crop Sci 48:85. https://doi.org/10.2135/cropsci2007.05.0287

Oliver KR, McComb JA, Greene WK (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol 5:1886–1901. https://doi.org/10.1093/gbe/evt141

Oppong A, Bedoya CA, Ewool MB et al (2014) Bulk genetic characterization of Ghanaian maize landraces using microsatellite markers. Maydica 59:1–8

Ortega Corona A, Guerrero Herrera MJ, Preciado Ortiz RE (eds) (2013) Diversidad y distribucion del maiz nativo y sus parientes silvestres en Mexico. Instituto Nacional de Investigaciones Forestales, Agricolas, y Pecuarias, Mexico, D.F.

Ortega Paczka R (2003) El maiz como cultivo II. In: Esteva G, Marielle C (eds) Sin Maiz No Hay Pais. Consejo Nacional para la Cultura y las Artes, Mexico, D.F., pp 123–154

Owens BF, Lipka AE, Magallanes-Lundback M et al (2014) A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. Genetics 198:1699–1716. https://doi.org/10.1534/genetics.114.169979

Peiffer JA, Romay MC, Gore MA et al (2014) The genetic architecture of maize height. Genetics 196:1337–1356. https://doi.org/10.1534/genetics.113.159152

Pflüger LA, Schlatter AR (1996) Isozyme variation in some races of maize from Argentina. Genet Resour Crop Evol 43:357–362. https://doi.org/10.1007/BF00132956

Piperno DR, Flannery KV (2001) The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. Proc Natl Acad Sci USA 98:2101–2103. https://doi.org/10.1073/pnas.98.4.2101

Pollak LM (2001) The History and success of the public-private project on germplasm enhancement of maize (GEM). Adv Agron 78:45–87

Pressoir G, Berthaud J (2004a) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92:88–94. https://doi.org/10.1038/sj.hdy.6800387

Pressoir G, Berthaud J (2004b) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. Heredity 92:95–101. https://doi.org/10.1038/sj.hdy.6800388

Provan J, Lawrence P, Young G et al (1999) Analysis of the genus *Zea* (*Poaceae*) using polymorphic chloroplast simple sequence repeats. Plant Syst Evol 218:245–256. https://doi.org/10.1007/BF01089230

Qi-Lun Y, Ping F, Ke-Cheng K, Guang-Tang P (2008) Genetic diversity based on SSR markers in maize (*Zea mays* L.) landraces from Wuling mountain region in China. J Genet 87:287–291. https://doi.org/10.1007/s12041-008-0046-y

Ramirez RE, Timothy DH, Efrain Diaz B, Grant UJ (1960) Races of maize in Bolivia. National Academy of Sciences—National Research, Washington, D. C

Realini MF, Poggio L, Cámara-Hernández J, González GE (2016) Intra-specific variation in genome size in maize: cytological and phenotypic correlates. AoB Plants 8:plv138. https://doi.org/10.1093/aobpla/plv138

Rebourg C, Chastanet M, Gouesnard B et al (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903. https://doi.org/10.1007/s00122-002-1140-9

Rebourg C, Gouesnard B, Charcosset A (2001) Large scale molecular analysis of traditional European maize populations. Relationships with morphological variation. Heredity 86:574–587. https://doi.org/10.1046/j.1365-2540.2001.00869.x

Reif JC, Hamrit S, Heckenberger M et al (2005) Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. Theor Appl Genet 111:906–913. https://doi.org/10.1007/s00122-005-0016-1

Reif JC, Warburton ML, Xia XC et al (2006) Grouping of accessions of Mexican races of maize revisited with SSR markers. Theor Appl Genet 113:177–185. https://doi.org/10.1007/s00122-006-0283-5

Remington DL, Thornsberry JM, Matsuoka Y et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484. https://doi.org/10.1073/pnas.201394398

Revilla P, Soengas P, Cartea ME et al (2003) Isozyme variability among European maize populations and the introduction of maize in Europe. Maydica 48:141–152

Revilla P, Soengas P, Malvar RA et al (1998) Isozyme variation and historical relationships among the maize races of Spain. Maydica 48:141–152

Romay MC, Millard MJ, Glaubitz JC, et al (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 14. https://doi.org/10.1186/gb-2013-14-6-r55

Ron Parra J, Hallauer AR (2010) Utilization of exotic maize germplasm. Plant breeding reviews. Wiley, Oxford, UK, pp 165–187

Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus Zea. Genetics 181:1399–1413. https://doi.org/10.1534/genetics.108.097238

Rubey L, Ward R, Tschirley D (1997) Maize research priorities. The role of consumer preferences.

In: Byerlee D, Eicher C (eds) Africa's emerging maize revolution. Lynne Rienner Publisher, Inc., Boulder, Colorado, pp 145–155

Ruiz Corral JA, Durán Puga N, Sánchez González J, Ron Parra J, González Eguiarte DR, Holland JB, Meidina García G (2008) Climatic adaptation and ecological descriptors of 42 Mexican maize races. Crop Sci 48:1502. https://doi.org/10.2135/cropsci2007.09.0518

Ruiz de Galarreta JI, Alvarez A (2001) Morphological classification of maize landraces from northern Spain. Genet Resour Crop Evol 48:391–400. https://doi.org/10.1023/A:1012074606561

Samayoa LF, Malvar RA, Moreno-González J, Ordás A, Revilla P (2016) Evaluation of white maize populations for quality and agronomic performance. Crop Sci 56:1173–1178. https://doi.org/10.2135/cropsci2015.08.0497

Sánchez DL, Liu S, Ibrahim R et al (2018) Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (*Zea mays* L.). Plant Sci 268:30–38. https://doi.org/10.1016/j.plantsci.2017.12.004

Sánchez JJ, Goodman MM (1992a) Relationships among the Mexican races of maize. Econ Bot 46:72–85. https://doi.org/10.1007/bf02985256

Sánchez JJ, Goodman MM (1992b) Relationships among Mexican and some North American and South American races of maize. Maydica 37:41–51

Sánchez JJ, Goodman MM, Bird RM, Stuber CW (2006) Isozyme and morphological variation in maize of five Andean countries. Maydica 51:25–42

Sánchez JJ, Goodman MM, Rawlings JO (1993) Appropriate characters for racial classification of maize. Econ Bot 47:44–59

Sánchez JJ, Goodman MM, Stuber CW (2000a) Isozymatic and morphological diversity in the races of maize of Mexico. Econ Bot 54:43–59. https://doi.org/10.1007/BF02866599

Sánchez JJ, Stuber CW, Goodman MM (2000b) Isozymatic diversity in races of maize of the Americas. Maydica 45:185–203

Sánchez JJ, Goodman MM, Stuber CW (2007) Racial diversity of maize in Brazil and adjacent areas. Maydica 52:13–30

Sanou J, Gouesnard B, Charrier A (1997) Isozyme variability in west African maize cultivars (Zea mays L.). May 42:1–11

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. https://doi.org/10.1126/science.1178534

Schön CC, Utz HF, Groh S et al (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485–498. https://doi.org/10.1534/GENETICS.167.1.485

Sharma L, Prasanna BM, Ramesh B (2010) Analysis of phenotypic and microsatellite-based diversity of maize landraces in India, especially from the North East Himalayan region. Genetica 138:619–631. https://doi.org/10.1007/s10709-010-9436-1

Simmonds NW (1993) Introgression and incorporation—strategies for the use of crop genetic resources. Biol Rev Camb Philos Soc 68:539–562. https://doi.org/10.1111/j.1469-185X.1993.tb01243.x

Smith JSC, Goodman MM, Stuber CW (1985) Genetic variability within U.S. maize germplasm. I. Historically important lines. Crop Sci 25:550–555. https://doi.org/10.2135/cropsci1985.0011183X002500030029x

Smith S (2007) Pedigree background changes in U.S. Hybrid maize between 1980 and 2004. Crop Sci 47:1914. https://doi.org/10.2135/cropsci2006.12.0763

Sood S, Flint-Garcia S, Wilcox MC, Holland JB (2014) Mining natural variation for maize improvement: selection on phenotypes and genes. In: Tuberosa R, Graner A, Frison E (eds) Genomics of Plant Genetic Resources. Springer, Dordrecht, pp 615–649

Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. Genetics 135

Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat Genet 43:1160–1163. https://doi.org/10.1038/ng.942

Sturtevant EL (1894) Notes on maize. Bull Torrey Bot Club 21: 319–343

Swarts K, Li H, Romero Navarro JA et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7:3. https://doi.org/10.3835/plantgenome2014.05.0023

Tallury SP, Goodman MM (1999) Experimental evaluation of the potential of tropical germplasm for temperate maize improvement. TAG Theor Appl Genet 98:54–61. https://doi.org/10.1007/s001220051039

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066. https://doi.org/10.1126/SCIENCE.277.5329.1063

Teixeira JEC, Weldekidan T, de Leon N et al (2015) Hallauer's Tusón: a decade of selection for tropical-to-temperate phenological adaptation in maize. Heredity 114:229–240. https://doi.org/10.1038/hdy.2014.90

Tenaillon MI, Sawkins MC, Long AD et al (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Tracy WF, Chandler MA (2008) The historical and biological basis of the concept of heterotic patterns in Corn Belt dent maize. In: Lamkey KR, Lee M (eds) Plant breeding: the Arnel R. Hallauer international symposium. Blackwell Publishing, Ames, Iowa, USA, pp 219–233

Uhr DV, Goodman MM (1995) Temperate maize inbreds derived from tropical germplasm: I. Testcross yield trials. Crop Sci 35:779. https://doi.org/10.2135/cropsci1995.0011183X003500030024x

Ullstrup AJ (1972) The impacts of the southern corn leaf blight epidemics of 1970–1971. Annu Rev Phytopathol 10:37–50. https://doi.org/10.1146/annurev.py.10.090172.000345

van Heerwaarden J (2007) Population genetics of traditionally managed maize: farming practice as a determinant of genetic structure and identity of maize landraces in Mexico. PhD thesis, Wageningen University

van Heerwaarden J, Doebley J, Briggs WH et al (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci 108:1088–1092. https://doi.org/10.1073/pnas.1013011108

van Heerwaarden J, Hufford MB, Ross-Ibarra J (2012) Historical genomics of North American maize. Proc Natl Acad Sci USA 109:12420–12425. https://doi.org/10.1073/pnas.1209275109

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299. https://doi.org/10.1007/s00122-009-1256-2

Vázquez-Carrillo G, García-Lara S, Salinas-Moreno Y, Bergvinson DJ, Palacios-Rojas N (2011) Grain and tortilla quality in landraces and improved maize grown in the highlands of Mexico. Plant Foods Hum Nutr 66:203–208. https://doi.org/10.1007/s11130-011-0231-7

Vega-Alvarez I, Santacruz-Varela A, Rocandio-Rodríguez M et al (2017) Genetic diversity and structure of native maize races from Northwestern Mexico. Pesqui Agropecuária Bras 52:1023–1032. https://doi.org/10.1590/s0100-204x2017001100008

Vigouroux Y, Glaubitz JC, Matsuoka Y et al (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95:1240–1253. https://doi.org/10.3732/ajb.0800097

Vigouroux Y, Mitchell S, Matsuoka Y et al (2005) An analysis of genetic diversity across the maize genome using microsatellites. Genetics 169:1617–1630. https://doi.org/10.1534/genetics.104.032086

Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. Proc Natl Acad Sci USA 103:17644–17649. https://doi.org/10.1073/pnas.0603080103

Warburton ML, Reif JC, Frisch M et al (2008) Genetic diversity in CIMMYT nontemperate maize germplasm: landraces, open pollinated varieties, and inbred lines. Crop Sci 48:617–624. https://doi.org/10.2135/cropsci2007.02.0103

Weatherwax P (1954) Indian corn in old America. Macmillan, New York

Weir BS, Cockerham CC (1984) Estimating F-statistics for analysis of population structure. Evolution 38:1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wellhausen EJ, Roberts LM, Hernandez E (1951) Razas de maíz en México: su orígen, caracteristicas y distribución. México. Secretaria de Agricultura y Ganadería. Oficina de Estudios Especiales/Fundación Rockefeller. México

Westengen OT, Berg PR, Kent MP, Brysting AK (2012) Spatial structure and climatic adaptation in African maize revealed by surveying SNP diversity in relation to global breeding and landrace panels. PLoS One 7: e47832. https://doi.org/10.1371/journal.pone.0047832

Whitlock MC (2008) Evolutionary inference from Q_{ST}. Mol Ecol 17:1885–1896. https://doi.org/10.1111/j.1365-294X.2008.03712.x

Whitt SR, Wilson LM, Tenaillon MI et al (2002) Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci USA 99:12959–12962

Wills DM, Fang Z, York AM et al (2017) Defining the role of the MADS-box gene, *Zea agamous-like1*, a target of selection during maize domestication. J Hered. https://doi.org/10.1093/jhered/esx073

Wills DM, Whipple CJ, Takuno S et al (2013) From many, one: genetic control of prolificacy during maize domestication. PLoS Genet 9:13. https://doi.org/10.1371/journal.pgen.1003604

Windhausen VS, Atlin GN, Hickey JM et al (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 Genes Genomes Genetics 2:1427–1436. https://doi.org/10.1534/g3.112.003699

Wright SI, Bi IV, Schroeder SG et al (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314. https://doi.org/10.1126/science.1107891

Yan J, Kandianis CB, Harjes CE et al (2010) Rare genetic variation at *Zea mays crtRB1* increases β-carotene in maize grain. Nat Genet 42:322–327. https://doi.org/10.1038/ng.551

Yang Q, He Y, Kabahuma M et al (2017) A gene encoding maize caffeoyl-CoA O-methyltransferase confers quantitative resistance to multiple pathogens. Nat Genet 49:1364–1372. https://doi.org/10.1038/ng.3919

Yang Q, Li Z, Li W et al (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc Natl Acad Sci 110:16969–16974. https://doi.org/10.1073/pnas.1310949110

Yao Q, Yang K, Pan G, Rong T (2007) Genetic diversity of maize (*Zea mays* L.) landraces from southwest China based on SSR data. J Genet Genomics 34:851–859. https://doi.org/10.1016/S1673-8527(07)60096-4

Yu X, Li X, Guo T et al (2016) Genomic prediction contributing to a promising global strategy to turbocharge gene banks. Nat Plants 2:16150. https://doi.org/10.1038/nplants.2016.150

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zila CT, Ogut F, Romay MC et al (2014) Genome-wide association study of Fusarium ear rot disease in the U.S.A. maize inbred line collection. BMC Plant Biol 14:372. https://doi.org/10.1186/s12870-014-0372-6

Toward Redesigning Hybrid Maize Breeding Through Genomics-Assisted Breeding

21

D. C. Kadam and A. J. Lorenz

Abstract

The pure line method of corn breeding proposed by George Shull in 1909 provided the foundation for contemporary hybrid breeding. This method involved development of inbreds by self-pollination and their subsequent evaluation in single-cross combinations. Over the past hundred years, this method underwent several modifications to efficiently generate homozygous inbred lines and identify superior hybrid combinations. Nevertheless, identification of promising pairs of lines that produce superior commercial hybrids is still challenging and constitutes the most expensive and critical operation in hybrid breeding. Advances in genotyping, predictive modeling, and computational capacities in recent years led to the development genome-based approaches for prediction of genetic values for complex traits. Simulation and experimental studies of genomic prediction of hybrid performance in maize have shown very promising results. With this background, we describe in this chapter a typical structure of a contemporary phenotypic hybrid maize breeding program; introduce the genomic selection (GS) approach and its application to a hybrid maize breeding pipeline; and review results of genomic prediction studies for hybrid maize, with particular emphasis on prediction of single crosses. We conclude with a discussion on future research needs and potential alternative hybrid breeding schemes in light of the rapidly developing field of genomics-assisted breeding.

21.1 Introduction

Breeding of hybrid crops was pioneered in maize and subsequently adopted in several other plant species of agronomic and horticultural importance. Hybrid cultivars provide higher yield compared to pure lines and open-pollinated varieties. Maize yield has increased by more than fivefold since the introduction of commercial hybrids in the 1930s (Pratt 2004; Kucharik and Ramankutty 2005; Assefa et al. 2017). Modern maize hybrids can outyield current open-pollinated varieties by 50–100% or more (Smith and Seiter 2003; Kutka et al. 2004; Smith and Cooper 2004). It has been estimated that about 50% of the increase in maize yield is due to genetic gain achieved through development of better hybrids (Tollenaar et al. 2000; Duvick 2005). Similarly, hybrid cultivars provide a yield advantage of about 20–30% in rice (Cheng et al. 2007), 10–25% in

D. C. Kadam · A. J. Lorenz (✉)
Department of Agronomy and Plant Genetics, University of Minnesota, Minneapolis, MN 55108, USA
e-mail: lore0149@umn.edu

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_21

wheat (Hoisington et al. 1999), and 35–40% in sorghum (Duvick and Cassman 1999). In addition to gain in yield, hybrids have other desirable features such as uniformity, yield stability, ease of combining multiple traits into one cultivar and continuous supply of good quality seed due to involvement of the private seed industry (Hallauer et al. 1988).

Shull (1908, 1909) first proposed the conceptual framework of hybrid maize breeding. He realized that an open-pollinated field of maize consisted of a complex array of hybrids and that self-fertilization would reduce the hybrids to "elementary species," or true breeding inbred lines in modern terminology. The true breeding lines could be combined again to reproduce the hybrid of highest genetic value to be found in Shull's abstract open-pollinated field of maize. Hence, Shull's suggested breeding method involved creation of pure lines by repeated generations of self-fertilization, followed by identification of superior hybrid combinations among the pure lines through testing. This method is referred to as the "pure line method of corn breeding." Although the pure line method of corn breeding still forms the basis of modern hybrid breeding, several modifications have occurred over the past hundred years to efficiently generate homozygous inbred lines and identify superior hybrid combinations between them. The main modifications include organization of inbred lines into heterotic groups to increase the probability of obtaining superior hybrids (Reif et al. 2005; Tracy and Chandler 2006); population improvement methods to increase the frequency of lines having good potential for hybrid performance (Comstock et al. 1949); doubled haploid (DH) technology to rapidly generate homozygous lines (Rober et al. 2005); and methods of early-generation selection outlined in the next paragraph.

Currently, heterotic groups are well established in temperate maize, and single-cross hybrids are exclusively made by crossing lines between heterotic groups. This greatly facilitates the efficient identification of superior hybrids, but the potential number of hybrid combinations still far exceeds the capacity of any phenotypic testing program. The development of DH technology has further exacerbated this dilemma. For example, even if the breeding program has just 100 inbreds from each heterotic group, the total number of hybrid combination to evaluate becomes 10,000. Most breeding programs will have far more than 100 inbred lines per heterotic group. Therefore, evaluation of lines for hybrid performance has been the most expensive and critical phase in hybrid maize breeding. Several approaches have been investigated to identify superior hybrids while circumventing the testing of all possible hybrid combinations. These approaches included inbred per se performance, topcross test (Jenkins and Brunson 1932), genetic distances based on molecular markers (Melchinger 1999; Lee et al. 2007), best linear unbiased prediction (BLUP) (Bernardo 1994, 1996), and markers associated with hybrid performance (Vuylsteke et al. 2000). These hybrid prediction approaches, however, have important limitations to their routine and effective use in hybrid breeding (Schrag et al. 2009).

In this chapter, we (1) describe a typical structure of a contemporary phenotypic hybrid maize breeding program; (2) introduce the genomic selection (GS) approach and its application to a hybrid maize breeding program; and (3) review results of genomic prediction studies for hybrid maize, with particular emphasis on prediction of single crosses. We conclude with a discussion on future research needs and potential alternative hybrid breeding schemes in light of the rapidly developing field of genomics-assisted breeding.

21.2 Structure of a Conventional Hybrid Maize Breeding Program

The process of commercial hybrid maize development consists of two stages: line development and hybrid evaluation (Fig. 21.1). In today's hybrid maize breeding programs, especially in the private sector, line development is most commonly accomplished through the creation of doubled haploid lines (DHLs) using haploid

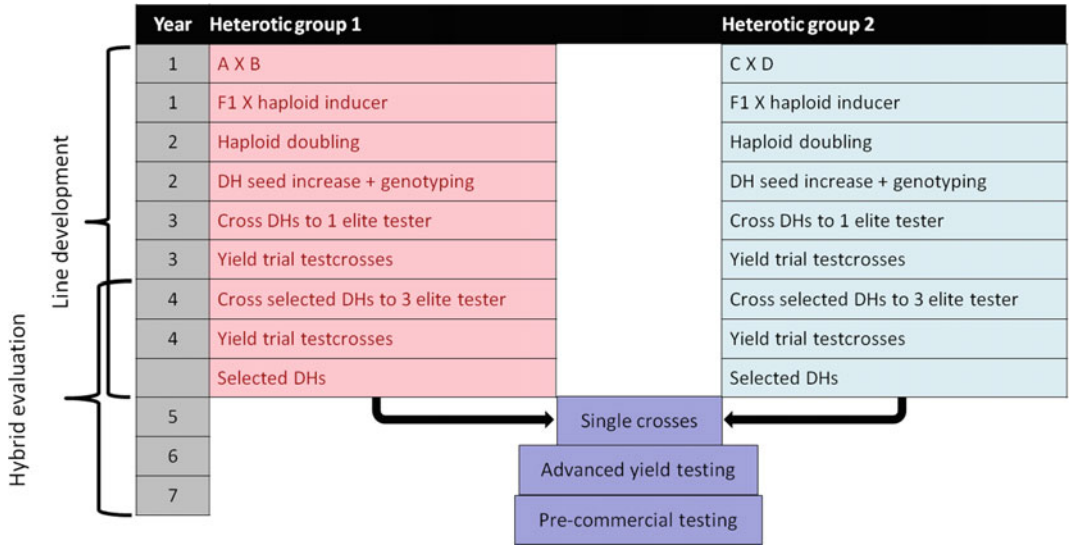

Fig. 21.1 Structure of a conventional hybrid maize development program illustrating the line development and hybrid evaluation stages based on doubled haploid lines. Breeding crosses are made within heterotic groups, and single crosses are made between heterotic groups. Multiple rows per year indicate multiple seasons enabled by use of winter nurseries

inducers and doubling agents (Prasanna et al. 2012; Prigge and Melchinger 2012). Breeding crosses to form new breeding populations are made by crossing elite lines within heterotic groups. Prior knowledge of performance of parental lines in earlier breeding cycles and pedigree relationships is used to determine the potential of specific crosses in terms of creating breeding populations with high average values and variances. Simulation studies indicate that selection of parents for crosses is far more important than number of crosses and number of lines derived from each cross (Bernardo 2003; Wegenast et al. 2008). Doubled haploid lines are typically produced from F1 plants instead of F2 plants to shorten the length of the breeding cycle (Longin et al. 2007). Recent studies, however, suggest deriving DHLs from F2 plants in order to increase the frequency of recombinants, preferably after selecting (within the F2 s) for highly heritable, additive traits such as some forms of disease and insect resistance (Wegenast et al. 2008; Bernardo 2009).

Characterization and selection of progeny lines involve sequential testing. Initially, lines are selected based on per se performance and topcross tests, while selections in the advanced stages are performed by general combining ability (GCA) and specific combining ability (SCA) evaluation in hybrid combinations. Selection based on per se performance is carried out for traits having reasonably high heritability and considered to be controlled by primarily additive effects, such as maturity, grain quality, and resistance to diseases and insects. In the case of RILs developed through pedigree breeding, topcross testing is commonly performed in the F4 and F5 generations after the lines having poor per se performance are discarded (Hallauer and Miranda 1988). Similarly, only those DHLs having suitable per se performance are evaluated in topcross test. Typically, two generations of topcross testing are conducted (Bernardo 2010; Geiger and Gordillo 2009; Heffner et al. 2010). With each round of topcross testing, the number of progeny lines advanced decreases while number of testers used increases. Narrow-based testers such as elite inbreds from the opposite heterotic group are commonly used for topcross testing. Theoretical and empirical results show that elite inbreds from opposite heterotic group generate topcross genetic variability as large as

when a poor performing tester is used (Hallauer and Lopez-Perez 1979; Bernardo 2010). Additionally, elite testers are more practical because superior hybrid combinations can be commercialized in a shorter period of time (Hallauer and Lopez-Perez 1979).

Lines selected based on topcross performance are further evaluated in many hybrid combinations to fully assess their GCA value and the SCA effects of specific single crosses. Resources are allocated to evaluate as many lines as possible at the topcross stage with intense selection, while at later stages, emphasis is placed on testing single crosses across many locations. The RILs or DHLs with superior GCA and stability identified from multi-location hybrid evaluation are often recycled as parents to develop source populations for line development (Smith 2004).

21.3 Genomic Selection

Molecular markers are widely used in modern maize breeding programs (Dudley et al. 1991; Prasanna and Hoisington 2003; Eathington et al. 2007; Bernardo 2008). The many uses of molecular markers to select upon simply inherited traits and perform traditional marker-assisted selection (MAS) or marker-assisted recurrent selection (MARS) for relatively complex traits are beyond the scope of this review. Our focus here is on genomic prediction and selection as it pertains to prediction of hybrid performance.

Genomic selection (GS) is defined as selection for a trait of interest using a set of molecular markers scored across the entire genome (Meuwissen et al. 2001). In MAS, only the markers identified as being significantly associated with the trait of interest using QTL mapping or other model selection algorithms are used for selection. Genomic prediction, on the other hand, uses all markers simultaneously to predict the genetic value of an individual; genomic selection simply denotes the selection of individuals on the basis of genomic predictions. Genomic selection has been shown to be more effective than MAS, especially for traits controlled by many small-effect QTLs. This is mainly because the separation of QTL detection and marker-effect estimation into two steps results in biased effect estimates, and the requirement of stringent statistical testing leaves many small-effect markers out of the prediction model and thus variation contributed by small-effect QTL are not captured in a MAS model (Jannink et al. 2010). Other difficulties of MAS, such as the use of specialized mapping populations for QTL identification, are avoided in a genomic prediction and selection scheme.

Genomic prediction and selection have become feasible in recent years because of the following advances in related sciences and technologies: 1. Efficient methods to genotype large numbers of markers (primarily single nucleotide polymorphisms; SNPs) (Thomson 2014); 2. Adaptation and development of better statistical methods to handle the high-dimensional marker data (Gianola et al. 2010; de los Campos et al. 2013); and 3. Availability of powerful statistical software packages (Butler et al. 2009; Endelman 2011; Pérez and de Los Campos 2014) and high capacity computational resources (Wu et al. 2011).

The central process of GS consists of two steps. The first step is the development of a genomic prediction model by combining genome-wide markers and phenotypic data on a subset of individuals called the training set (TRS). In the second step, the genomic prediction model is used to calculate genomic estimated breeding values (GEBVs) for individuals comprising the target set which have only genotypic data and no phenotypic data, and subsequently selections are made on the basis of genomic predictions in order to advance lines to the next stage of phenotypic evaluation or for use as parents to form new breeding populations (Fig. 21.2). Assessment of prediction accuracy is an important consideration for GS implementation, as well as for determining the various factors under control of the breeder that affect prediction accuracy. Ideally, the accuracy of genomic predictions is the correlation between true breeding value (TBV) and GEBV. However, in practice, TBV is unknown. Considering

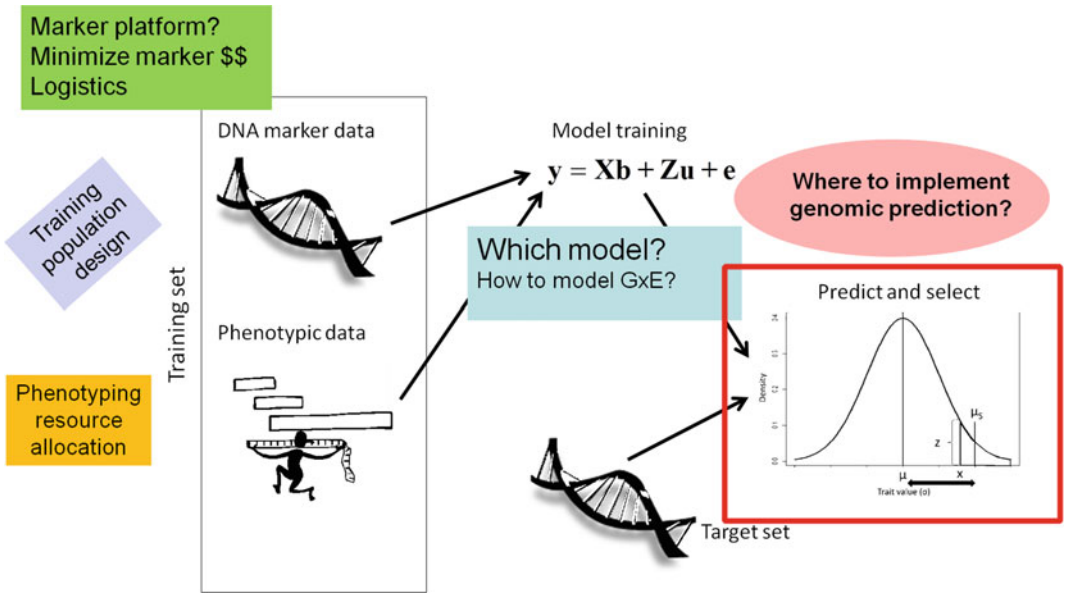

Fig. 21.2 Schematic of genomic prediction and selection, and factors related to its optimization at each stage

$$\mathrm{cor}(\hat{g}, \mathrm{y}) = \mathrm{cor}(\hat{g}, g) \times \mathrm{cor}(g, y)$$

where y is a vector of phenotypes; $\hat{g}$ is a GEBV; and g is TBV, the ***prediction accuracy*** is estimated as cor ($\hat{g}$, y)/h where h is the square root of heritability on an entry-mean basis (Legarra et al. 2008; Hayes et al. 2009). ***Predictive ability***, often defined as cor ($\hat{g}$, y), is also commonly used for these purposes, which is valid as long as it is recognized that predictive abilities are expected to be biased downwards because h is always less than one. The best way to calculate prediction accuracy or predictive ability is by using an appropriate cross-validation scheme, with the exact scheme depending on the goals and questions of the researcher.

21.3.1 Factors Affecting Genomic Prediction Accuracy

Genomic prediction accuracy is affected by many factors under the control of the breeder, as well as other factors outside the breeder's control, including choice of statistical model for genomic prediction, genetic relationships between the TRS and target population, trait heritability, TRS size, and marker density (Fig. 21.2).

21.3.1.1 Statistical Model

The primary challenge in building a genomic prediction model is that the number of molecular markers (p) (i.e., predictors) is typically larger than the number of individuals (n) in the TRS (i.e., observations). This is known as the "large p and small n" problem (de los Campos et al. 2013). Under a least-squares estimation framework, this leads to a situation where either not enough degrees of freedom exist for estimation of marker effects or, if $n > p$ yet p is comparatively large, a high degree of multicollinearity among marker effects results in large variances around the estimates.

To confront this problem, a slew of alternative statistical models have been deployed with different underlying assumptions (de los Campos et al. 2013; González-Recio et al. 2014). These models can be broadly separated into two categories: parametric and nonparametric. Briefly, parametric models assume a certain form of relationship between genetic value and marker covariates. Marker effects are estimated either

using shrinkage or a combination of shrinkage and variable selection. Commonly used parametric models include ridge regression best linear unbiased prediction (RRBLUP), genomic best linear unbiased prediction (GBLUP), and Bayesian models. In RRBLUP, marker effects are assumed to be random and normally distributed with a common variance, resulting in equal shrinkage of their effects. The consequence of equal shrinkage is that markers having large effect on the trait are more underestimated compared to those with smaller effect. Thus, the RRBLUP model is more appropriate when there are few or no large-effect QTL and many small-effect QTLs, which is the case with most quantitative traits. Genomic best linear unbiased prediction uses a genomic relationship matrix (GRM) calculated from marker genotypes instead of calculating individual marker effects (VanRaden 2008). Under the assumption of multivariate normality, GBLUP has been shown to be mathematically equivalent to RRBLUP (Habier et al. 2007), although slight differences in realized prediction accuracies between GBLUP and RRBLUP could be attributed to marker-QTL linkage disequilibrium (LD). In contrast to RRBLUP, many Bayesian models are able to relax the assumption of common marker-effect variances and allow marker-specific variances, effectively allowing unequal shrinkage of marker effects (de los Campos et al. 2013).

The prediction accuracy of different parametric GS models depends upon the genetic architecture of the trait and LD structure in the population (Lorenz et al. 2011; Lin et al. 2014). Simulation results indicate that RRBLUP and GBLUP rely strongly on kinship, while Bayesian models focus more on LD between marker and QTL than on kinship (Habier et al. 2007; Zhong et al. 2009). Thus, if some large-effect QTL controls an important amount of variation for a trait, Bayesian models can provide better accuracy than RRBLUP and GBLUP. Alternatively, if trait variation is dominated by small-effect QTL, both models can achieve similar prediction accuracies. Results of empirical studies, however, have shown comparable performance of both types of models across different types of trait architectures (Lorenzana and Bernardo 2009; Moser et al. 2009). When strong long-range LD exists in a population, the effects of major QTLs can be captured by markers well apart from the QTL, resulting in good prediction accuracies of RRBLUP and GBLUP.

Nonparametric models take a different approach by not making strong assumptions about the form of relationship between marker covariates and genetic value (González-Recio et al. 2014). Instead, these models seek the form that best fits the TRS data while maintaining some generality for new data. In other words, their main focus is on prediction. These models, therefore, hold potential to capture non-additive effects without explicitly modeling them and, hence, can provide better prediction of phenotypes for complex traits where non-additive effects are important (Gianola et al. 2006, 2010). Some commonly used nonparametric GS models include reproducing kernel Hilbert spaces (RKHS), support vector regression (SVR), and neural network (NN) (González-Recio et al. 2014). When the non-additive gene effects are important for a given trait, simulation and some experimental studies indicate better performance of nonparametric models over parametric models (Heslot et al. 2012; Pérez-Rodríguez et al. 2012; Howard et al. 2014; Jiang and Reif 2015).

21.3.1.2 Genetic Relationship

The genetic relationship between the TRS and the target population is the most important factor influencing the accuracy of genomic prediction. The TRS needs to be representative of the target population in order to obtain good prediction accuracy. The closer genetic relationship benefits the prediction accuracy in at least two ways: (1) It reduces the effective population size, generating strong long-range LD between marker and QTL that is consistent in phase between TRS and target populations; (2) Marker-by-genetic background effects, to the extent that they exist, are ameliorated by training and predicting within similar genetic backgrounds. The prediction accuracy is, therefore, expected to be highest for training and prediction within a family, followed

by families connected by one shared parent. Empirical GS studies in many crops, including maize, have stressed the importance of genetic relationship for obtaining good prediction accuracy (Albrecht et al. 2011, 2014; Riedelsheimer et al. 2013; Jacobson et al. 2014; Lorenz and Smith 2015).

21.3.1.3 Heritability

Heritability is an important determinant of achievable prediction accuracy. High heritability enables accurate estimation of marker effects because phenotypic variation is mostly composed of genetic variation with only little confounding effect of environmental factors. A highly significant correlation has been observed between heritability and prediction accuracy in empirical studies in maize (Lorenzana and Bernardo 2009; Jacobson et al. 2014). Although the accuracy of both GS and phenotypic selection is affected by heritability, GS becomes more efficient over phenotypic selection with a decrease in heritability (Bernardo and Yu 2007; Viana et al. 2016). The genetic relationship information and LD between markers and QTLs enable GS to outperform the phenotypic selection under low heritability situations.

21.3.1.4 Training Set Size

Increasing the TRS size allows a more accurate estimation of marker effects and consequently enhances the prediction accuracy. A positive correlation between TRS size and prediction accuracy has been reported from studies in maize (Lorenzana and Bernardo 2009; Albrecht et al. 2011; Zhao et al. 2012). It is important to note that increasing the genetic relationship between TRS and target population is a more effective way to increase the prediction accuracy than increasing the TRS size by adding less related individuals (Riedelsheimer et al. 2013). However, a reasonable TRS size is required to obtain reliable prediction even with a close genetic relationship between TRS and target population (Schulz-Streeck et al. 2012).

21.3.1.5 Number and Type of Markers

The number of markers required to obtain optimal prediction accuracy depends on LD in the population under consideration. If the LD is high, fewer markers are required and vice versa. Due to the availability of cheap and abundant genome-wide SNPs in recent years (Davey et al. 2011), marker density should not be a limiting factor to obtain maximum achievable prediction accuracy. Marker number can be considerably increased by efficient marker imputation methods (Huang et al. 2014; Jacobson et al. 2015; Technow and Gerke 2017). The type of markers used can also influence the prediction accuracy. Solberg et al. (2008) reported that three times higher SNP density is required to obtain prediction accuracies comparable to simple sequence repeat (SSR) markers, because SSRs have multiple alleles and therefore contain more information. The multi-allelic system of SSRs can be mimicked by constructing haplotype containing multiple SNPs, but this still requires a considerably larger number of SNPs in the initial dataset. The improvement in prediction accuracy using haplotypes is minimal, however, especially when SNP density is high (Calus et al. 2008). In another study, Poland et al. (2012) found greater GS accuracy using SNPs obtained from genotyping by sequencing (GBS) than diversity array technology (DArT) marker. The prospects of transcript and metabolite data are also currently investigated for their predictive potential (Xu et al. 2017).

21.3.2 Considerations Related to GS Implementation that Can Affect Prediction Accuracy

In addition to the above-mentioned factors which affect the genomic prediction accuracy, there are some other relevant issues for implementation of genomic prediction in a plant breeding program. These include TRS optimization, the effect of

population size vs. number of replications, GS considering multiple traits simultaneously, and modeling genotype × environment interaction. These aspects are briefly described below.

21.3.2.1 Training Set Composition

One of the key issues in implementing GS is how to design a TRS that maximizes prediction accuracy with minimum resources spent on phenotyping and genotyping. When using existing phenotypic data, estimating the achievable prediction accuracy would be greatly beneficial. Deterministic formulae to estimate the achievable GS accuracy from available data were developed by Daetwyler et al. (2008, 2010). These formulae, however, have limited utility because of they require an estimate of the effective number of chromosome segments (M_e) for a trait in a population, which is a value not always easily estimable (Combs and Bernardo 2013; Lian et al. 2014). Newer methods of predicting achievable GS accuracy that do not require estimates of M_e should be developed and validated.

When phenotypic data are not already available or not sufficient to obtain required prediction accuracy, an important question is how to select the most informative individuals for phenotyping and model training. Rincent et al. (2012) tested a TRS optimization algorithm using the prediction error variance (PEV) and coefficient of determination (CD) as objective functions, as well as stratified sampling in two diversity maize panels consisting of 300 dent and 300 flint lines. The TRS selected based on PEV and CD objective functions provided higher reliability than randomly selected TRS for different sizes. Further, the usefulness of PEV and CD mean criteria was demonstrated by Isidro et al. (2015) using different datasets.

21.3.2.2 Population Size Versus Number of Replications

Phenotyping expenses involved in GS model training could also be reduced by optimal allocation of resources between number of individuals phenotyped and number of replications. In a simulation experiment, Lorenz (2013) found a great flexibility for resource allocation between number of individuals and replication in a GS context. However, in another simulation study, Zhong et al. (2009) reported a benefit of increasing population size over number of replications when predictions were performed several generations away from the TRS population. One explanation for these contradictory results could be the different population structure in these two studies. Zhong et al. (2009) used several biparental populations which can generate tight linkage between marker and QTL, whereas Lorenz (2013) used a single biparental population. In the case of predicting average genotypic performance across multiple environments, Endelman et al. (2014) reported an advantage of phenotyping a larger number of genotypes with limited replication over testing a smaller number of genotypes with greater replication. Further simulation and empirical studies on resource allocation between population size and replication might shed light on this issue.

21.3.2.3 Multiple-Trait Genomic Selection

Most of the GS studies to date targeted a single trait (i.e., single-trait GS). In practice, however, selection is often performed for several traits simultaneously. Genomic selection considering multiple traits at a time (i.e., multiple-trait GS) is currently under investigation. Multiple traits can be targeted in GS directly by using a selection index formed by weighting the traits by their importance and summing across traits for each individual, or indirectly by specifying the variance–covariance structure among the traits in the model (Schulthess et al. 2016). Multiple-trait GS would be advantageous over single-trait GS under the following situations: The primary trait has low heritability, while the secondary trait has high heritability; the phenotypic data are only partially available for the primary trait and completely available for secondary trait; or if there is strong genetic correlation between primary and secondary traits (Jia and Jannink 2012; Guo et al. 2014). These situations are expected to be less common in practice. Therefore, the accuracies of multiple-trait GS are similar or only marginally higher than single-trait GS in

empirical studies in maize (dos Santos et al. 2016; Lyra et al. 2017) as well as other crops (Jia and Jannink 2012; Rutkoski et al. 2012; Bao et al. 2015; Schulthess et al. 2016).

21.3.2.4 Genotype-by-Environment Interaction

The single environment model cannot exploit information on performance of genotypes in other environments. In the case of multi-environment models, information across environment is borrowed at the expense of forcing the marker effects to be constant across environments. Therefore, the need of multi-environment GS models, which enables borrowing of information across environments, has been recognized. Three different ways to accommodate genotype-by-environment interaction in GS models have recently been proposed, which include (1) use of covariance structure (Burgueño et al. 2012), (2) use of explicit environmental covariates (Jarquín et al. 2014), and (3) modeling marker-by-environment interaction effects (Lopez-Cruz et al. 2015). The accuracies of these multi-environment GS models were substantially higher than a single environment model, especially for predicting the performance of genotypes in environments in which genotypes were not tested. These findings underline the importance of considering genotype-by-environment interaction in GS models. Further comparisons of these three forms of multi-environment GS models would be desirable to identify a suitable model for routine implementation. One advantage of a marker-by-environment interaction model is that it can shed light on which genomic regions are most responsible for genotype-by-environment interaction (Lopez-Cruz et al. 2015). Similarly, multi-environment GS models using environmental covariates offer an opportunity to enhance the statistical model with biological knowledge. In a further extension of this approach, integration of crop growth models (CGMs) in genomic prediction is currently seen as an important area of research for prediction of complex trait phenotypes (Bustos-Korts et al. 2016).

21.4 Genomic Prediction Accuracy in Maize

Several studies have examined the potential of GS at different stages of hybrid maize breeding, including inbred per se performance, topcross performance, and single-cross performance (Table 21.1). GS for per se performance (yield) of inbred lines has comparatively limited scope as the value of a line in hybrid breeding is determined by its performance in hybrid combinations. However, GS can be beneficial for traits such as disease resistance, which are often phenotyped on a line per se basis because of their largely additive genetic architecture. Technow et al. (2013) investigated the accuracies of genomic prediction of northern corn leaf blight resistance among inbreds belonging to dent and flint heterotic groups. Prediction accuracies were low to moderate. They found considerable benefit from increasing the training set size within heterotic groups from N = 25 to N = 75, as well as by combining inbreds across two heterotic groups into the same TRS. Riedelsheimer et al. (2013) evaluated the prospects of combining multiple differently related populations into a TRS for predicting per se performance of lines for five traits including Gibberella ear rot severity and three kernel yield component traits. They observed a considerable decline in predictive ability (from 0.59 to 0.25) when full-sib lines were replaced by half-sib lines, but predictive abilities were improved (from 0.36 to 0.39) when half-sib lines were available from both the parents instead of only one parent of the target population. A negative effect of combining unrelated populations into the TRS was also observed.

Topcross testing (cross of inbred lines with a common tester inbred line) in a maize breeding program primarily serves the purpose of identifying candidate inbred progenies with superior GCA which should be subjected to additional rounds of more intensive evaluation, both in terms of crossing to more testers and testing at more locations. Genomic prediction of topcross performance, therefore, could aid in saving field

Table 21.1 Summary of published studies on genomic selection for per se performance, topcross performance, and single-cross performance

Reference	Brief description	Experimental material	Model	Cross-validation	Prediction accuracy[d]
A. *Genomic selection for per se performance*					
Technow et al. (2013)	Accessed the prospects of genomic prediction of northern corn leaf blight resistance and combining inbred lines across heterotic groups into TRS	*Germplasm*: 100 dent and 97 flint inbred lines *Markers*: 37908 SNPs	GBLUP	CV_WW[a] CV_AW[b] CV_AA[c]	0.33–0.64 (CV_WW) 0.08–0.3 (CV_AW) 0.37–0.71 (CV_AA)
Riedelsheimer et al. (2013)	Investigated the effect of different level of relatedness between TRS and TS on prediction accuracy within BP for two disease traits and three grain yield component traits	*Germplasm*: 635 DH lines from the five interconnected BP *Markers*: 16741 SNPs	GBLUP	CV_WW CV_AW	0.59 (CV_WW), 0.05–0.34 (CV_AW)
B. *Genomic selection for topcross performance*					
Lorenzana and Bernardo (2009)	Compared the prediction accuracies of MLR, GBLUP, and e-Bayes methods and evaluated the effect of TRS size and number of markers	*Germplasm*: Testcrosses of RIL/DHLs belonging to three BP *Markers*: 1339 SSR or RFLP; 125 SNPs	GBLUP e-Bayes	CV_WW	0.25–0.64
Albrecht et al. (2011)	Examined the accuracies of within versus across family prediction. Also assessed the effect of TRS size and different approaches of estimating genetic relationship	*Germplasm*: Testcrosses of 1380 DH lines from 36 BP belonging to dent heterotic group *Markers*: 1152 SNPs	GBLUP	CV_WW CV_AW CV_AA	0.26–0.59 (CV_WW) 0.47–0.48 (CV_AW) 0.72–0.74 (CV_AA)
Riedelsheimer et al. (2012)	Investigated the usefulness of genome and metabolite-based prediction	*Germplasm*: testcrosses of 285 diverse inbred lines *Markers*: 56110 SNPs and 130 metabolites	RRBLUP	CV_AA	0.60–0.78
Schulz-Streeck et al. (2012)	Evaluated the advantage of modeling main and population-specific marker effects. Also compared RRBLUP, ridge regression, LASSO, and elastic net	*Germplasm*: Testcrosses of 312 DH lines from five BP *Markers*: 39339 SNPs	RRBLUP RR LASSO EN	CV_AW CV_AA	0.024–0.31 (CV_AW) 0.28–0.37 (CV_AA) Note: predictive ability (heritability not given)
Windhausen et al. (2012)	Evaluated the prospects of marker effects estimated in diversity panel for prediction within a biparental population	*Germplasm*: Testcrosses of 255 inbreds from diversity panel and 150 inbreds belonging to 5 BP *Markers*: 18695 SNPs	GBLUP	CV_AW CV_AA CV_AW_{group} CV_AA_{group}	−0.42 to 0.37 (CV_AW) 0.46–0.54 (CV_AA) 0.14–0.26 (CV_AW_{group}) 0.15–0.39 (CV_AA_{group})

(continued)

Table 21.1 (continued)

Reference	Brief description	Experimental material	Model	Cross-validation	Prediction accuracy[d]
Zhao et al. (2012)	Compared the prediction within and across biparental families. Also, evaluated the effect of modeling preselected markers with low genetic background interaction effect	*Germplasm*: Testcrosses 788 $F_{3:4}$ lines from six BP *Markers*: 960 SNPs	GBLUP	CV_WW CV_AW CV_AA	0.40–0.64 (CV_WW) 0.39–0.70 (CV_AW) 0.45–0.69 (CV_AA)
Crossa et al. (2013)	Compared the different methods of incorporating genotyping by sequencing (GBS) marker data for genomic prediction with GBLUP and RKHS	*Germplasm*: Testcrosses 505 DH lines and diverse panel of 296 maize inbred lines *Markers*: GBS	GBLUP RKHS	CV_AA	0.60–0.90
Massman et al. (2013b)	Assessed the usefulness of marker effects estimated from single-cross data for testcross prediction	*Germplasm*: Testcrosses of 5 BP along with 479 single crosses between 59 BSSS inbreds and 49 NSSS inbreds Markers: 669 SNPs	GBLUP RRBLUP	CV_AW	−0.08 to 0.36
Albrecht et al. (2014)	Accessed the efficiency of prediction across genetic groups and tester. Also compared the potential of predicting across locations and across years	*Germplasm*: Testcrosses of 1,073 and 857 DH lines derived from multiple biparental families *Markers*: 56110 SNPs	GBLUP	CV_WW_{group} CV_AW_{group} $CV\_AW_{group/tester}$ CV_AA_{group}	0.36–0.77 (CV_WW_{group}) 0.31–0.35 (CV_AW_{group}) 0.14–0.53 ($CV\_AW_{group/tester}$) 0.45–0.74 (CV_AA_{group})
Jacobson et al. (2014)	Evaluated the usefulness of GCA model for genome-wide selection within a BP	*Germplasm*: Testcrosses of 970 BP *Markers*: 49 to 100 SNPs	RRBLUP	CV_AW CV_WW	−0.16 to 0.63 (CV_WW) 0.02–0.65 (CV_AW)
C. *Genomic selection for single-cross performance*					
Bernardo (1994)	First used BLUP for prediction of single-cross performance	*Germplasm*: 54 single crosses between six BSSS inbreds and nine NSSS inbreds *Markers*: RFLP	GBLUP	Random sampling	0.63–0.80
Maenhout et al. (2007)	Compared SVR and GBLUP for prediction of single-cross performance	*Germplasm*: 2371 single crosses between 105 BSSS and 93 Iodent lines *Markers*: 75 SSR and AFLP	SVR GBLUP	LOOCV	0.66

(continued)

Table 21.1 (continued)

Reference	Brief description	Experimental material	Model	Cross-validation	Prediction accuracy[d]
Maenhout et al. (2010)	Compared single-cross prediction accuracies of SVR and GBLUP in unbalanced dataset. Also, compared prediction accuracy obtained through cross-validation and field evaluation	*Germplasm*: 2354 single crosses between 105 BSSS and 92 Iodent lines *Markers*: 75 SSR and AFLP	SVR GBLUP	*k*-fold CV for T1 and T0 single crosses	0.62–0.78 (T1) 0.32–0.58 (T0)
Massman et al. (2013a)	Compared BLUP with RRBLUP for single-cross prediction	*Germplasm*: 479 single crosses between 59 BSSS inbreds and 49 NSSS inbreds *Markers*: 669 SNPs	BLUP RRBLUP	*k*-fold CV for T2, T1 single crosses	0.87 (T2) 0.73–0.75 (T1)
Technow et al. (2012)	In a simulation study, investigated the effects of marker density, convergent or divergent parental populations, number of tested parents, genetic model and estimation method on the prediction accuracy	*Germplasm*: 10000 in silico single crosses between 100 dent and 100 flint inbreds *Markers*: 39627 SNPs	GBLUP BayesB	Random sampling for T2, T1 and T0 single crosses	0.84–0.91 (T2) 0.74–0.84 (T1) 0.65–0.76 (T0)
Technow et al. (2014)	Evaluated the prospects of single-cross prediction using GBLUP and BayesB	*Germplasm*: 1254 single crosses between 123 dent and 86 flint inbred lines *Markers*: 35478 SNPs	GBLUP BayesB	*k*-fold CV for T2, T1, and T0 single crosses	0.86–0.92 (T2) 0.82–0.86(T1) 0.75–0.78(T0)
Kadam et al. (2016)	Examined the potential of genomic prediction of early-stage single crosses	*Germplasm*: 312 single crosses between inbreds belonging to dent and flint heterotic groups *Markers*: 2296 SNPs	GBLUP	LOOCV for T2, T1, and T0 single crosses	0.67–0.76 (T2) 0.37–0.63 (T1) 0.28–0.40 (T0)
Zenke-Philippi et al. (2016)	Compared accuracies of genomic and transcriptomic data for prediction of single crosses	*Germplasm*: 98 single crosses between seven flint and 14 dent lines *Markers*: 970 AFLP and 10810 expression profiles	RRBLUP	Random sampling for T2 and T0 single crosses	0.70–0.72 (T2) 0.47–0.49 (T0)
Westhues et al. (2017)	Compared accuracies of genomic, transcriptomic and metabolomic data for prediction of single crosses	*Germplasm*: 1536 single crosses between 142 dent and 103 flint lines *Markers*: 21565 SNPs, 1323 expression profiles and 375 metabolite	GBLUP	Random sampling for T0 single crosses	~ 0.22–0.77 (T0)

(continued)

Table 21.1 (continued)

Reference	Brief description	Experimental material	Model	Cross-validation	Prediction accuracy[d]
Schrag et al. (2018)	Compared accuracies of genomic, transcriptomic (including s-RNA) and metabolomic data for prediction of single crosses	*Germplasm*: 1567 single crosses between 143 dent and 104 flint lines *Markers*: 37392 SNPs, 300 mRNA, 10736 s-RNA expression profiles and 148 metabolite	GBLUP	Random sampling for T0 single crosses	~0.75–0.85 (T0)

Different cross-validation scenarios: [a]Training set (TRS) and test set (TS) sampled within a biparental population; [b]TRS sampled across biparental populations and TS sampled within a biparental population; [c]TRS and TS sampled across biparental populations. Subscripts group and group/tester are used to denote above three cross-validation scenarios with reference to group and group/tester instead of biparental population

[d]Prediction accuracies for grain yield (unless specified) when TRS and TS were evaluated in the same environment/s or across environments

Acronyms: BP—Biparental populations; GCA—General combining ability; BSSS—Iowa stiff stalk synthetic; LOOCV—Leave-one-out cross-validation; NSSS—Non-stiff stalk synthetic; e-Bayes—empirical Bayes; GBLUP—Genomic best linear unbiased prediction; MLR—Multiple linear regression; RRBLUP—Ridge regression best linear unbiased prediction; RKHS—Reproducing kernel Hilbert space

testing resources and overall costs if genotyping is less expensive than topcrossing and phenotyping. The genomic prediction studies for topcross performance have looked at the effect of different factors such as TRS size, marker density, prediction within versus across populations, and prediction across testers. As expected, topcross prediction accuracy is generally benefited by increasing marker number and TRS size (Lorenzana and Bernardo 2009; Albrecht et al. 2011; Zhao et al. 2012). These studies indicated that a minimum TRS size of about 50–100 when predicting within a biparental population (family) and about 300–400 when predicting for populations related by at least one common parent (half-sib) are required to obtain prediction accuracy above 0.5, assuming moderate to high heritability. Also, about 100 markers for GS within a biparental population and 200–400 markers for GS with multiple interconnected populations are suggested to obtain optimal prediction accuracy. The mean topcross prediction accuracies within biparental population were shown to be moderate to high (Lorenzana and Bernardo 2009; Albrecht et al. 2011; Zhao et al. 2012).

A disadvantage of genomic prediction within a biparental population is the need to phenotype a subset of individuals from the same population which increases the time and cost. Also, individual population sizes need to be sufficiently large to reliably perform within-population predictions (Schulz-Streeck et al. 2012). It would therefore be advantageous if performance of lines within a biparental population could be predicted before the population is phenotyped. In this context, several studies investigated the effect of estimating marker effects across populations to predict within each population (Albrecht et al. 2011; Zhao et al. 2012; Jacobson et al. 2014). The prediction accuracies were similar or slightly lower than within-population prediction when the topcross information of half-sib lines from both the parents were available. The prediction accuracies were severely decreased when topcross information of half-sib lines from only one or none of the parent were available. Furthermore, when a diversity panel is used to estimate marker effects, prediction accuracies were very poor (Windhausen et al. 2012). Few possible reasons for decrease in accuracy of genomic prediction for the across versus within-population scenario include marker x population interaction, epistasis and different linkage phases between marker and QTL among populations (Schulz-Streeck et al. 2012). In an effort to enhance prediction accuracy, models including

population-specific marker effects (Schulz-Streeck et al. 2012) or only the preselected markers having low marker x genetic background interaction (Zhao et al. 2012) were investigated. However, no improvement in prediction accuracy was observed. In a different scenario, when estimation and prediction were performed across the biparental populations, the prediction accuracies were higher compared to within biparental population prediction (Albrecht et al. 2011; Schulz-Streeck et al. 2012; Zhao et al. 2012; Windhausen et al. 2012). The increase in prediction accuracy resulted from differences in mean performances of populations rather than kinship between estimation and prediction set and LD between markers and QTLs (Windhausen et al. 2012). As genetic variation among populations can be efficiently exploited through parental selection, GS application is not needed in this scenario.

21.5 Comparison of Phenotypic, Marker-Assisted Recurrent Selection, and Genomic Selection in Maize

Despite the theoretical advantages of genomic selection over phenotypic and marker-assisted selection, application of genomic selection requires empirical results showing its effectiveness. Before we delve deeper into the topic of single-cross prediction, a brief review of studies showing genetic gains in maize through GS is provided.

The relative efficiencies of phenotypic selection, MARS, and GS have been compared in simulation and field studies in maize. Bernardo and Yu (2007) first showed the relative effectiveness of GS in plant breeding through simulation. They simulated MARS and GS for testcross performance using DHLs derived from a single biparental population. Previous research has thoroughly shown the advantage of MARS over phenotypic selection in terms of genetic gain per unit time (reviewed by Bernardo (2008)), and hence, GS was not directly compared to phenotypic selection in this study. After two cycles of GS or MARS, response to selection from GS was 18 to 43% larger than that of MARS across different numbers of QTLs and levels of heritability. In a follow-up study, Massman et al. (2013b) provided the first published field results comparing GS to MARS. Their experiment involved two cycles of GS and MARS for testcross performance for stover and yield indices in a population consisting of 233 RILs derived from B73 and Mo17. The realized gains were 14–50% larger with GS compared to MARS. Beyene et al. (2015) compared the genetic gain for grain yield in eight biparental populations under managed drought stress conditions using GS versus a pedigree-based phenotypic selection scheme. Because the length of phenotypic selection cycles is longer than GS cycles, three cycles of GS were compared to one cycle of phenotypic selection. The average gain from GS per cycle across eight populations was 0.086 mg ha^{-1}. Hybrids derived from cycle 3 produced 7.3% higher grain yield than those developed by the pedigree breeding scheme. Recently, Vivek et al. (2017) reported a study on genetic gain under drought conditions using phenotypic selection and GS in two biparental populations. Cycle 1 was formed by intermating the top 10% families selected based on testcross performance. Subsequently, the second cycle of selection was conducted based on phenotypic selection of inbred progeny per se performance yielding C2-per se PS and GS using a model trained on testcross performance yielding C2-GS. Testcrosses of C2-GS showed, 11% higher grain yield than those of C2-per se PS averaged across populations and growing conditions. The authors estimated that gains in grain yield per year of GS were 6–85% more than per se PS.

In another recent study, Zhang et al. (2017) first applied rapid cycle GS to a multi-parental population derived from ten elite maize parents for four recombination cycles. While they did not directly compare their results to PS or MARS, the realized genetic gain with GS cycles (C1 to C4) was 0.225 mg ha^{-1} $cycle^{-1}$, equivalent to 0.100 mg ha^{-1} $year^{-1}$. The authors determined that the gain per year achieved using GS was equivalent to previous reports in tropical maize

for PS, and that the shorter cycle time of GS and reduced costs of genotyping as compared to phenotyping provide a strong advantage for GS in maize breeding. These results, combined with the foregoing comparisons between GS and MARS, show the effectiveness of GS for making genetic gain in maize, and support incorporation of GS into applied maize breeding programs.

21.6 Genomic Prediction of Single-Cross Hybrids in Maize

As discussed in the introduction of this chapter, determination of which single crosses to test has been an important consideration for maize breeders since the dawn of hybrid breeding over a century ago. Just as genomic prediction is revolutionizing plant breeding programs in choosing parents and allocating resources for field testing, it also holds potential for changing the scope of genetic evaluation. Rather than narrowing the field of candidate inbred lines to test in single-cross combinations (Fig. 21.1), genomic prediction allows the in silico evaluation of all possible single-cross combinations given all the candidate inbred lines have been genotyped (Fig. 21.3).

The use of GBLUP for predicting single-cross hybrids in maize goes back to the introduction of BLUP methodology to the plant breeding literature in the 1990s (Bernardo 1994). Bernardo (1994) used restriction fragment length polymorphisms (RFLPs) to estimate covariances among GCA effects for a small number of inbred lines, as well as covariances of SCA effects for the corresponding single crosses. The data were divided into tested and untested single crosses, and correlations between predicted yield and observed yield demonstrated that predictions were highly accurate, with correlations ranging from 0.69 to 0.80. Covariances estimated using RFLP data produced predictions that were slightly more accurate than those based on pedigree data. Similar results were found in a similar study by Charcosset et al. (1998).

Since these earlier studies, prospects for genomic prediction of single-cross performance in maize have been further investigated using larger populations and much larger marker sets. The streamlined use of markers in modern breeding programs and the intense interest in incorporating GS into all breeding decisions have increased interest in the use of single-cross genomic prediction. Remarkably, all of these studies have reported very high prediction accuracies of single-cross performance using cross-validation, ranging from about 0.70 to over 0.90, even with modestly sized training populations. Beyond reporting high prediction accuracies and indicating great promise for this application of GS, these studies have explored factors affecting prediction accuracy of single crosses such as statistical model choice, estimation of dominance or SCA effects, and TRN composition. Given the uniqueness of this application of genomic prediction, findings related to how these factors affect accuracy will be specifically discussed in the context of single-cross prediction.

Following the original single-cross prediction studies of Bernardo (1994, 1996), most studies have focused on the use of the GBLUP model. Indeed, GBLUP has proven difficult to beat in terms of prediction accuracy. Maenhout et al. (2007, 2010) compared ε-insensitive support vector machine regression (ε-SVR) to GBLUP and found that GBLUP was slightly more accurate under various prediction scenarios. In a simulation study, Technow et al. (2012) found a BayesB model to be significantly better than GBLUP, but only slightly (advantage of $\sim$0.02 in prediction accuracy). Using field data on grain yield, however, the same authors found no difference between BayesB and GBLUP in a follow-up study (Technow et al. 2014). Another twist to modeling genome-wide marker effects is fitting them as population-specific (i.e., heterotic group) or unspecific to a population (Technow et al. 2012). In their simulation study, Technow et al. (2012) found that population-specific effects provided a slight advantage, especially under situations of low marker-QTL LD where it

is more likely that marker-QTL linkage phases are switched between divergent populations.

Single-cross performance is a function of the GCA effects of the parents and the SCA effect of the specific parental combination. Although, the proportion of SCA/GCA variance among inter-heterotic group single crosses is generally expected to be low (Reif et al. 2007), some published studies have reported significant proportions of SCA/GCA variance ranging between 0.1 and 0.2 among single crosses between dent and flint inbreds (Technow et al. 2014), and Iowa stiff stalk synthetic (BSSS) and non-stiff stalk (NSSS) inbreds. Therefore, many single-cross prediction studies in maize have evaluated the benefit of predicting SCA effects and using them in addition to GCA effects. Both parametric (Schrag et al. 2006; Technow et al. 2012, 2014; Kadam et al. 2016) and nonparametric (Maenhout et al. 2007, 2010; Kadam and Lorenz 2018) models have been used for SCA effect prediction. A simulation study by Technow et al. (2012) showed subtle benefits to estimating and including SCA effects in genomic prediction models, especially for the case of "convergent" heterotic groups where marker alleles associated with QTL effects were simulated to be at similar allele frequency in the two heterotic groups. This created a situation where SCA variance was 25% of the total genotypic variance. However, little to no benefit of estimating and including SCA effects was observed when "divergent" heterotic groups were simulated with QTL markers at very different allele frequencies. In studies using real data, virtually no benefit to including SCA effects in genomic prediction models has been found (Kadam et al. 2016; Bernardo 1994; Schrag et al. 2018; Westhues et al. 2017; Maenhout et al. 2010; Schrag et al. 2006). These studies all used inter-heterotic-group single crosses, and thus, the relative proportions of SCA variance were found to be low, and GCA variance was the predominant type of genotypic variance. Also, in those cases where SCA variance is relatively important, SCA effects have proven to be difficult to predict, even using nonparametric models (Kadam and Lorenz 2018).

Another important objective of single-cross prediction studies has been to investigate the effect of number of tested parents (i.e., parents of the single crosses in the target set were already evaluated as parents for other single crosses in the TRS) of a single crosses on prediction accuracy. In that context, genomic prediction of T2 (both parents tested), T1 (either male or female parent tested), and T0 (neither parent tested) single crosses were performed. The number of tested parents had greater influence on single-cross prediction accuracy than other factors such as TRS size, number of markers, and the statistical model (Technow et al. 2012, 2014; Kadam et al. 2016). Genomic prediction accuracies were found to be highest for T2 single crosses, followed by T1 and T0 single crosses (Technow et al. 2012, 2014; Massman et al. 2013a, b; Kadam et al. 2016). The highest accuracy of T2 single crosses demonstrates the close relationship between TRS and TS because both the parents of single crosses contained in the target set are also tested among the single crosses forming the TRS. Similarly, intermediate and low accuracies for prediction of T1 and T0 single crosses were due to decreasing relationships between TRS and target set in these prediction scenarios. The increase in accuracy when increasing the number of tested parents can also be explained by an increase in identical allele copies between TRS and TS (Technow et al. 2012).

Beyond genomic data, modern "-omics" technologies provide the opportunity to collect large amounts of data from various biological strata, which could be useful for predictive modeling of complex traits through capturing biological interactions across these strata (Frisch et al. 2010; Zenke-Philippi et al. 2016b; Westhues et al. 2017; Schrag et al. 2018). Interestingly, some recently published studies have reported enhanced prediction accuracies (up to 14% improvement) using transcriptomic data in a GBLUP framework (Schrag et al. 2018). This was especially the case for T0 single crosses, which is an important result because T0 single crosses make up the majority of single crosses

needing to be predicted. These authors also modeled metabolomic and small RNA data, but found no benefit compared to genomic data. Many questions related to these applications are outstanding, such as statistical model and cost-to-benefit ratio. Also, because the transcriptome is sensitive to the environment and varies across tissues, questions related to the development stage of tissue sampling, type of tissue to be sampled, and growth conditions of plants being sampled are likely very important variables when using this type of data for predictive purposes.

21.7 Summary and Future Directions

Much of the history of maize hybrid breeding methodology research has focused on strategies to efficiently identify superior single-cross hybrids. Concepts such as heterotic groups and combining ability have been combined with methods such as topcross testing and BLUP to streamline hybrid breeding programs and maximize the probability of identifying those superior single crosses. Despite these advances, the ability to evaluate all possible single crosses between all inbred lines created from the very earliest stages of a breeding cycle has been hindered by the finite resources of a breeding program, and the seemingly infinite resources needed to tackle this huge problem. The wide-scale adoption and integration of genomic prediction hold great potential to close this gap and return hybrid maize breeding to the conceptual framework envisioned by Shull (1908). That is, the evaluation of all possible single crosses to find that most superior, individual genotype in Shull's abstract open-pollinated field of maize. From a cost standpoint, this ideal scenario is possible because the cost of genotyping inbred lines from each heterotic group is only a multiple of 2N, while the cost of phenotyping all possibly hybrids is a multiple of N^2, where N is the number of lines per heterotic group. At N = 1000, $(2N)/N^2 \approx 0.002$; this ratio decreases as N increases. The numbers are clear from a cost standpoint, but the successful use of genomic prediction for evaluating single crosses comes down to prediction accuracy. Fortunately, studies that have reported genomic prediction accuracies of single crosses have all reported remarkably high accuracies, ranging from 0.70 to greater than 0.90. This has even been found to be true for single crosses made between inbred lines from a small number of biparental families, representing the earlier stages of a hybrid maize breeding cycle (Kadam et al. 2016; Kadam and Lorenz 2018). Because of these high accuracies, we, along with other authors (e.g., Technow et al. 2014; Westhues et al. 2017), believe that maize breeders should begin to consider foregoing topcross testing to save time (Fig. 21.3), save resources, and possibly avoid discarding those inbred lines that happen to just interact negatively with the one tester used in the early stages of a breeding cycle, but still could

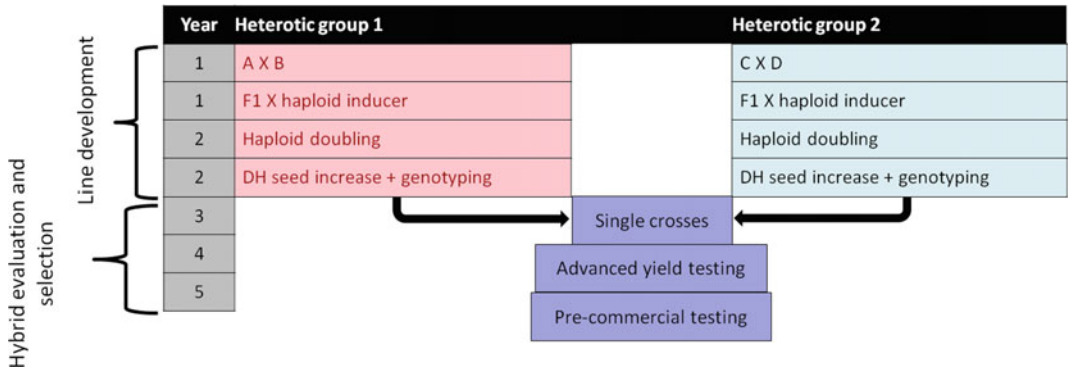

Fig. 21.3 Structure of a hybrid maize breeding scheme that aggressively uses genomic prediction of single crosses to replace topcross testing

combine well with other lines. We recognize that complete reliance on prediction during these early stages may be met with reluctance. At the very least, maize breeders should consider calculating genomic predictions for all possible single crosses alongside their topcrossing program and then advance those lines included as parents of single crosses of high predicted value to field testing if they were not advanced by topcross testing.

One drawback of this methodology is the need to form a training population with many representative single crosses. This could require a large amount of manual labor for hand pollinations to make seed for the individual single crosses. After all, a main impetus behind topcross testing is the ease of making hybrid seed, requiring only the de-tasseling of the female lines being topcrossed. To help ameliorate this, a research area that needs further examination is the design of such training populations for predicting single-cross performance. Using marker data to identify which single crosses would be most informative to a genomic prediction model could reduce training population size while maintaining accuracy. Additional areas of research we feel are important include accurate prediction of SCA effects and prediction of T0 single crosses. While prediction of T0 single crosses has been met with some success, these prediction accuracies are lower than the T2 and T1 cases. Devising models to more accurately predict SCA effects, possibly using data from different biological strata as in Schrag et al. (2018), could help the prediction of T0 single crosses as well.

In conclusion, genomic prediction has opened a new chapter for the design of hybrid maize breeding programs. The ability to predict using inexpensive and easily obtained predictor variables (i.e., markers) opens up a universe of possibilities and extends the scope of evaluation for the breeder from what was once limited to the genotypes realized, to all possible genotypes, both realized and yet to be realized. This new scope will alter many approaches to selective plant breeding, especially hybrid maize breeding.

Acknowledgements We thank Amritpal Singh for the helpful discussion during the preparation of outline of this chapter. Dnyaneshwar C. Kadam was supported by Indian Council of Agricultural Research International Fellowship (ICAR-IF).

References

Albrecht T, Auinger H-J, Wimmer V et al (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theor Appl Genet 127:1375–1386. https://doi.org/10.1007/s00122-014-2305-z

Albrecht T, Wimmer V, Auinger H-J et al (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350. https://doi.org/10.1007/s00122-011-1587-7

Assefa Y, Prasad PV, Carter P et al (2017) A new insight into corn yield: trends from 1987 through 2015. Crop Sci 57:2799–2811

Bao Y, Kurle JE, Anderson G, Young ND (2015) Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Mol Breed 35:128

Bernardo R (2003) Parental selection, number of breeding populations, and size of each population in inbred development. Theor Appl Genet 107:1252–1256

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Bernardo R (2009) Should maize doubled haploids be induced among F1 or F2 plants? Theor Appl Genet 119:255–262

Bernardo R (2010) Breeding for quantitative traits in plants. Stemma Press

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

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

Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna BM, Banziger M, Crossa J (2015) Genetic gains in grain yield through genomic selection in eight bi-parental Maize populations under drought stress. Crop Sci 55(1):154. https://doi.org/10.2135/cropsci2014.07.0460

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719

Bustos-Korts D, Malosetti M, Chapman S, van Eeuwijk F (2016) Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics. Crop systems biology. Springer, pp 55–82

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R reference manual. State Qld Dep Prim Ind Fish Brisb

Calus MPL, De Roos APW, Veerkamp RF et al (2008) Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561

Charcosset A, Bonnisseau B, Touchebeuf O et al (1998) Prediction of maize hybrid silage performance using marker data: comparison of several models for specific combining ability. Crop Sci 38:38–44

Cheng S-H, Zhuang J-Y, Fan Y-Y et al (2007) Progress in research and development on hybrid rice: a super-domesticate in China. Ann Bot 100:959–966

Combs E, Bernardo R (2013) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. Plant Genome 6

Comstock RE, Robinson HF, Harvey PH (1949) Breeding procedure designed to make maximum use of both general and specific combining ability. Agron J

Crossa J, Beyene Y, Kassa S et al (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes Genomes Genet 3:1903–1926

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185: 1021–1031

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:e3395

Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345. https://doi.org/10.1534/genetics.112.143313

dos Santos JPR, de Castro Vasconcellos RC, Pires LPM et al (2016) Inclusion of dominance effects in the multivariate GBLUP model. PLoS ONE 11:e0152045

Dudley JW, Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31:718–723

Duvick DN (2005) The contribution of breeding to yield advances in maize (Zea mays L.). Adv Agron 86:83–145

Duvick DN, Cassman KG (1999) Post–green revolution trends in yield potential of temperate maize in the North-Central United States. Crop Sci 39:1622–1630

Eathington SR, Crosbie TM, Edwards MD et al (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S–154

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255

Endelman JB, Atlin GN, Beyene Y et al (2014) Optimal design of preliminary yield trials with genome-wide markers. Crop Sci 54:48–59

Frisch M, Thiemann A, Fu J et al (2010) Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor Appl Genet 120:441–450

Geiger HH, Gordillo GA (2009) Doubled haploids in hybrid maize breeding. Maydica 54:485–499

Gianola D, de los Campos G, González-Recio O et al (2010) Statistical learning methods for genome-based analysis of quantitative traits. In: Proceedings of the 9th world congress on genetics applied to livestock production. Gesellschaft für Tierzuchtwissenschaften eV, pp 1–6

Gianola D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173:1761–1776

González-Recio O, Rosa GJ, Gianola D (2014) Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livest Sci 166:217–231

Guo G, Zhao F, Wang Y et al (2014) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:30

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397

Hallauer AR, Lopez-Perez E (1979) Comparisons among testers for evaluating lines of corn. In: Proceedings of the annual corn and sorghum research conference (USA)

Hallauer AR, Miranda JB (1988) Quantitative genetics in plant breeding. Iowa State

Hallauer AR, Russell WA, Lamkey KR (1988) Corn breeding

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50:1681–1690

Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52:146–160

Hoisington D, Khairallah M, Reeves T et al (1999) Plant genetic resources: what can they contribute toward increased crop productivity? Proc Natl Acad Sci 96:5937–5943

Howard R, Carriquiry AL, Beavis WD (2014) Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. G3 Genes Genomes Genet 4:1027–1046

Huang BE, Raghavan C, Mauleon R et al (2014) Efficient imputation of missing markers in low-coverage genotyping-by-sequencing data from multiparental crosses. Genetics 197:401–404

Isidro J, Jannink J-L, Akdemir D et al (2015) Training set optimization under population structure in genomic selection. Theor Appl Genet 128:145–158

Jacobson A, Lian L, Zhong S, Bernardo R (2014) General combining ability model for genomewide selection in a biparental cross. Crop Sci 54:895–905

Jacobson A, Lian L, Zhong S, Bernardo R (2015) Marker imputation before genomewide selection in biparental maize populations. Plant Genome 8

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Jarquín D, Crossa J, Lacaze X et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607

Jenkins MT, Brunson AM (1932) Methods of estimating inbred lines of maize in crossbred combinations

Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522

Jiang Y, Reif JC (2015) Modeling epistasis in genomic selection. Genetics 201:759–768

Kadam DC, Potts SM, Bohn MO et al (2016) Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. G3 Genes Genomes Genet 6:3443–3453

Kadam DC, Lorenz AJ (2018) Evaluation of nonparametric models for genomic prediction of early-stage single crosses in maize. Crop Sci (Accepted)

Kucharik CJ, Ramankutty N (2005) Trends and variability in US corn yields over the twentieth century. Earth Interact 9:1–29

Kutka F, Conway P, Christensen J (2004) The heritage maize project (FNC00-301) study of open-pollinated corn: final report to the North Central region sustainable agriculture research and education program. Herit Maize Proj FNC 00-301 Study Open-Pollinated Corn Final Rep North Cent Reg Sustain Agric Res Educ Program NCR-SARE Program Univ Neb Linc

Lee EA, Ash MJ, Good B (2007) Re-examining the relationship between degree of relatedness, genetic effects, and heterosis in maize. Crop Sci 47:629–635

Legarra A, Robert-Granié C, Manfredi E, Elsen J-M (2008) Performance of genomic selection in mice. Genetics 180:611–618

Lian L, Jacobson A, Zhong S, Bernardo R (2014) Genomewide prediction accuracy within 969 maize biparental populations. Crop Sci 54:1514–1522

Lin Z, Hayes BJ, Daetwyler HD (2014) Genomic selection in crops, trees and forages: a review. Crop Pasture Sci 65:1177–1191

Longin CFH, Utz HF, Reif JC et al (2007) Hybrid maize breeding with doubled haploids: III. Efficiency of early testing prior to doubled haploid production in two-stage selection for testcross performance. Theor Appl Genet 115:519–527

Lopez-Cruz M, Crossa J, Bonnett D et al (2015) Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. Genes Genomes Genet G3:g3–114

Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3 Genes Genomes Genet 3:481–491

Lorenz AJ, Chao S, Asoro FG et al (2011) Genomic selection in plant breeding: knowledge and prospects. Advances in agronomy. Elsevier, pp 77–123

Lorenz AJ, Smith KP (2015) Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci 55:2657–2667

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Lyra DH, de Freitas Mendonça L, Galli G et al (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. Mol Breed 37:80

Maenhout S, De Baets B, Haesaert G (2010) Prediction of maize single-cross hybrid performance: support vector machine regression versus best linear prediction. Theor Appl Genet 120:415–427

Maenhout S, De Baets B, Haesaert G, Van Bockstaele E (2007) Support vector machine regression for the prediction of maize hybrid performance. Theor Appl Genet 115:1003–1013

Massman JM, Gordillo A, Lorenzana RE, Bernardo R (2013a) Genomewide predictions from maize single-cross data. Theor Appl Genet 126:13–22. https://doi.org/10.1007/s00122-012-1955-y

Massman JM, Jung H-JG, Bernardo R (2013b) Genome-wide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci 53:58–66

Melchinger AE (1999) Genetic diversity and heterosis. Genet Exploit Heterosis Crops 99–118

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Moser G, Tier B, Crump RE et al (2009) A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet Sel Evol 41:56

Pérez P, de Los Campos G (2014) Genome-wide regression & prediction with the BGLR statistical package. Genetics 114

Pérez-Rodríguez P, Gianola D, González-Camacho JM et al (2012) Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3 Genes Genomes Genet 2:1595–1605

Poland J, Endelman J, Dawson J et al (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113

Prasanna BM, Chaikam V, Mahuku G (2012) Doubled haploid technology in maize breeding: theory and practice. CIMMYT

Prasanna BM, Hoisington D (2003) Molecular breeding for maize improvement: an overview

Pratt RC (2004) An historical examination of the development and adoption of hybrid corn: a case study in Ohio. Maydica 49:155–172
Prigge V, Melchinger AE (2012) Production of haploids and doubled haploids in maize. Plant cell culture protocols. Springer, pp 161–172
Reif JC, Hallauer AR, Melchinger AE (2005) Heterosis and heterotic patterns in maize. Maydica 50:215
Reif JC, Gumpert FM, Fischer S, Melchinger AE (2007) Impact of interpopulation divergence on additive and dominance variance in hybrid populations. Genetics 176(3):1931–1934. https://doi.org/10.1534/genetics.107.074146
Riedelsheimer C, Czedik-Eysenberg A, Grieder C et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220
Riedelsheimer C, Endelman JB, Stange M et al (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194:493–503
Rincent R, Laloë D, Nicolas S et al (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (Zea mays L.). Genetics 192:715–728
Rober FK, Gordillo GA, Geiger HH (2005) In vivo haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica 50:275
Rutkoski J, Benson J, Jia Y et al (2012) Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. Plant Genome 5:51–61
Schrag TA, Frish M, Dhillon BS, Melchinger AE (2009) Marker-based prediction of hybrid performance in maize single-crosses involving doubled haploids. Maydica 54:353
Schrag TA, Melchinger AE, Sørensen AP, Frisch M (2006) Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. Theor Appl Genet 113(6):1037–1047. https://doi.org/10.1007/s00122-006-0363-6
Schrag TA, Westhues M, Schipprack W et al (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. Genetics 208:1373–1385
Schulthess AW, Wang Y, Miedaner T et al (2016) Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor Appl Genet 129:273–287
Schulz-Streeck T, Ogutu JO, Karaman Z et al (2012) Genomic selection using multiple populations. Crop Sci 52:2453–2461
Shull GH (1908) The composition of a field of maize
Shull GH (1909) A pure-line method in corn breeding. J Hered 51–58
Smith CW (2004) Corn: origin, history, technology, and production. Wiley
Smith D, Cooper M (2004) Changes in performance, parentage, and genetic diversity of successful corn hybrids, 1930–2000. Corn Orig Hist Technol Prod 4:65
Smith M, Seiter S, Mt. Pleasant J, Kutka F (2003) Performance of open-pollinated corn varieties for grain and silage production. In: Proceedings of the 58th Northeast corn improvement conference
Solberg TR, Sonesson AK, Woolliams JA, others (2008) Genomic selection using different marker types and densities. J Anim Sci 86:2447–2454
Technow F, Bürger A, Melchinger AE (2013) Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. G3 Genes Genomes Genet 3:197–203
Technow F, Gerke J (2017) Parent-progeny imputation from pooled samples for cost-efficient genotyping in plant breeding. PLoS ONE 12:e0190271
Technow F, Riedelsheimer C, Schrag TA, Melchinger AE (2012) Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor Appl Genet 125:1181–1194
Technow F, Schrag TA, Schipprack W et al (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197:1343–1355. https://doi.org/10.1534/genetics.114.165860
Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. Plant Breed Biotechnol 2:195–212
Tollenaar M, Ying J, Duvick DN (2000) Genetic gain in corn hybrids from the northern and central corn belt. In: Proceedings of the 55th corn and sorghum research conference. American Seed Trade Association Washington, DC, pp 53–62
Tracy WF, Chandler MA (2006) The historical and biological basis of the concept of heterotic patterns in corn belt dent maize. In: Proceedings of plant breeding: the Arnel R. Hallauer international symposium. Wiley, pp 219–233
VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423
Viana JMS, Piepho H-P et al (2016) Quantitative genetics theory for genomic selection and efficiency of breeding value prediction in open-pollinated populations. Sci Agric 73:243–251
Vivek BS, Krishna GK, Vengadessan V et al (2017) Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. Plant Genome 10
Vuylsteke M, Kuiper M, Stam P (2000) Chromosomal regions involved in hybrid performance and heterosis: their AFLP®-based identification and practical use in prediction models. Heredity 85:208–218
Wegenast T, Longin CFH, Utz HF et al (2008) Hybrid maize breeding with doubled haploids. IV. Number versus size of crosses and importance of parental selection in two-stage selection for testcross performance. Theor Appl Genet 117:251–260

Westhues M, Schrag TA, Heuer C et al (2017) Omics-based hybrid prediction in maize. Theor Appl Genet 130:1927–1939

Windhausen VS, Atlin GN, Hickey JM et al (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. Genes Genomes Genet 2:1427–1436. https://doi.org/10.1534/g3.112.003699

Wu X-L, Beissinger TM, Bauck S et al (2011) A primer on high-throughput computing for genomic selection. Front Genet 2

Xu Y, Xu C, Xu S (2017) Prediction and association mapping of agronomic traits in maize using multiple omic data. Heredity 119:174

Zenke-Philippi C, Thiemann A, Seifert F et al (2016) Prediction of hybrid performance in maize with a ridge regression model employed to DNA markers and mRNA transcription profiles. BMC Genom 17:262

Zhang X, Pérez-Rodríguez P, Burgueño J et al (2017) Rapid cycling genomic selection in a multi-parental tropical maize population. Genes Genomes Genet G3: g3–117

Zhao Y, Gowda M, Liu W et al (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776

Zhong S, Dekkers JC, Fernando RL, Jannink J-L (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355–364

Correction to: The Maize Genome

Jeffrey Bennetzen, Sherry Flint-Garcia, Candice Hirsch and Roberto Tuberosa

Correction to:
J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes, https://doi.org/10.1007/978-3-319-97427-9

The originally published version of chapter 12 (The Genetics and Genomics of Virus Resistance in Maize) and chapter 20 (Harnessing maize biodiversity) did not acknowledge that "Author name" is an employee of the U.S government and that therefore the copyright of the work belongs to the "U.S government". This has been corrected in the updated version.

The updated version of these chapters can be found at
https://doi.org/10.1007/978-3-319-97427-9_12
https://doi.org/10.1007/978-3-319-97427-9_20

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_22