Theoretical and Applied Genetics (2019) 132:817–849
https://doi.org/10.1007/s00122-019-03306-3

REVIEW ARTICLE

Technological advances in maize breeding: past, present and future

Carson Andorf[1] · William D. Beavis[2] · Matthew Hufford[3] · Stephen Smith[2] · Walter P. Suza[2] · Kan Wang[2] · Margaret Woodhouse[1] · Jianming Yu[2] · Thomas Lübberstedt[2]

Received: 15 August 2018 / Accepted: 5 February 2019 / Published online: 23 February 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Maize has for many decades been both one of the most important crops worldwide and one of the primary genetic model organisms. More recently, maize breeding has been impacted by rapid technological advances in sequencing and genotyping technology, transformation including genome editing, doubled haploid technology, parallelled by progress in data sciences and the development of novel breeding approaches utilizing genomic information. Herein, we report on past, current and future developments relevant for maize breeding with regard to (1) genome analysis, (2) germplasm diversity characterization and utilization, (3) manipulation of genetic diversity by transformation and genome editing, (4) inbred line development and hybrid seed production, (5) understanding and prediction of hybrid performance, (6) breeding methodology and (7) synthesis of opportunities and challenges for future maize breeding.

Introduction

Maize (*Zea mays* L.) has become adapted "to the broadest range of climatic conditions of all crops, from 40S in Chile to 50N in Canada and Russia, from sea level in the West Indies to elevations above 3400 m in the Andes" (Bouchet et al. 2013). Global maize agriculture was significantly enabled through adaptation to temperate environments that initially occurred during a 2000-year period following introduction into the North American continent ca. 4000 years before present (BP) (Bouchet et al. 2013; Swarts et al. 2017). Record global maize grain production of 1054 million metric tonnes was achieved during 2016/17, rising by approx. 15 mio. metric tonnes/year since 1961. Eighty-five percent of maize grain is produced by nine countries and the European Union. China and the USA collectively accounted for 56% of global maize production in 2017/18 (National Corn Growers Association 2018). In addition, global production of silage maize increased > fourfold since 1961 to 18 million tonnes contributed by a doubling of area under production to 1.4 million ha and a > twofold increase in yield to 12.8 tonnes/ha (FAOSTAT 2018). Population growth and greater demand for animal products, particularly in developing countries, continue to increase demands for maize usage as food, feed and fuel (ethanol) and other industrial raw materials.

Maize occupies approximately equal areas of production in the tropics and temperate environments, yet the majority (70%) of maize production occurs under temperate conditions (Edmeades et al. 2017). Most global production is provided by hybrid maize (Duvick 2005a, b; Masuka et al. 2017a, b). Hybrids developed by CIMMYT yield > 20% more than OPVs under optimal conditions, and the disparity is magnified to 30–> 60% under abiotic and biotic stress conditions (Masuka et al. 2017a). However, open-pollinated varieties (OPVs) provide the majority of seed supply in some regions provided by the formal breeding sector (e.g., West Africa), albeit with much regional variation (Kassie et al. 2012) and due to many cited issues including seed supply (Pixley 2006; Gaffney et al. 2016). More resources in terms of breeding support over a longer period of time have been directed toward maize improvement in temperate climates than have been applied, to date, to the improvement in maize production in the tropics (Edmeades et al. 2017) and heterotic patterns are not firmly established in tropical maize populations (Betran et al. 2003; Reif et al. 2003; Wen et al. 2011). Nonetheless, mean rates of yield

Communicated by Rajeev K. Varshney.

✉ Thomas Lübberstedt
thomasL@iastate.edu

1 USDA-ARS, Ames, IA 50011-1010, USA

2 Department of Agronomy, Iowa State University, Agronomy Hall, Ames, IA 50011-1010, USA

3 Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011-1010, USA

Inari Exhibit 1015
Inari v. Pioneer

increase in tropical environments are now similar to those in temperate climates at 74–75 kg/ha/year, i.e., at 1% or 2.3% in temperate and tropical environments, respectively. However, there are major disparities between South America and SE Asia (128–142 kg/ha/year) and sub-Saharan Africa, Central America and the Caribbean (27–40 kg/ha/year) (Edmeades et al. 2017). Figure 1 shows trends in maize production and yields for countries or regions that collectively provide ca. 70% of global maize grain production. Global trends in maize yields during 1961–2008 are shown in Fig. 2 (Ray et al. 2012). Low rates of yield gain are usually due to several interrelated factors that can change with circumstances and which include a lack of uptake of improved varieties, poor soil fertility, weeds, pests, disease and droughts (Fischer et al. 2014).

Comparing the contribution of breeding (genetic gain) to increased yields among different studies is challenging due to specific contextual issues including germplasm, breeding strategies, cultivation and harvesting practices, initial yield levels, methodology and length of study. Genetic gain data achieved using single-cross hybrids for eight countries with rates of gain ranging from 50 to 194 kg/ha/year (Smith et al. 2014). Stagnating or declining maize yields are not due to a lack of potential genetic gain. For example, results from CIMMYT hybrid maize breeding program for eastern and southern Africa demonstrate rates of genetic gain of 109.4 kg/ha/year (optimal conditions), 12 kg/ha/year (low *N*), 23–32 kg/ha/year (drought) and 141.3 kg/ha/year under maize streak virus infestation (Masuka et al. 2017b). Accelerated adoption of improved drought-tolerant maize varieties could generate from US$362–US$590 during 7 years and use of low-*N*-tolerant varieties has similar financial gross benefits including US$100–US$136 in benefits to producers (Masuka et al. 2017a, b). Strategies to improve maize productivity in China include higher seed quality and the development of hybrids that preserve individual plant yield under higher planting densities (Ci et al. 2011; Li et al. 2011; Niu et al. 2013; Qin et al. 2016).

Rates of production have increased since 1991 in Iowa, Illinois and Minnesota, while remaining stable for Indiana and decreasing in Nebraska (Fig. 3). The rate of genetic gain for maize (Smith et al. 2014), updated to include hybrids

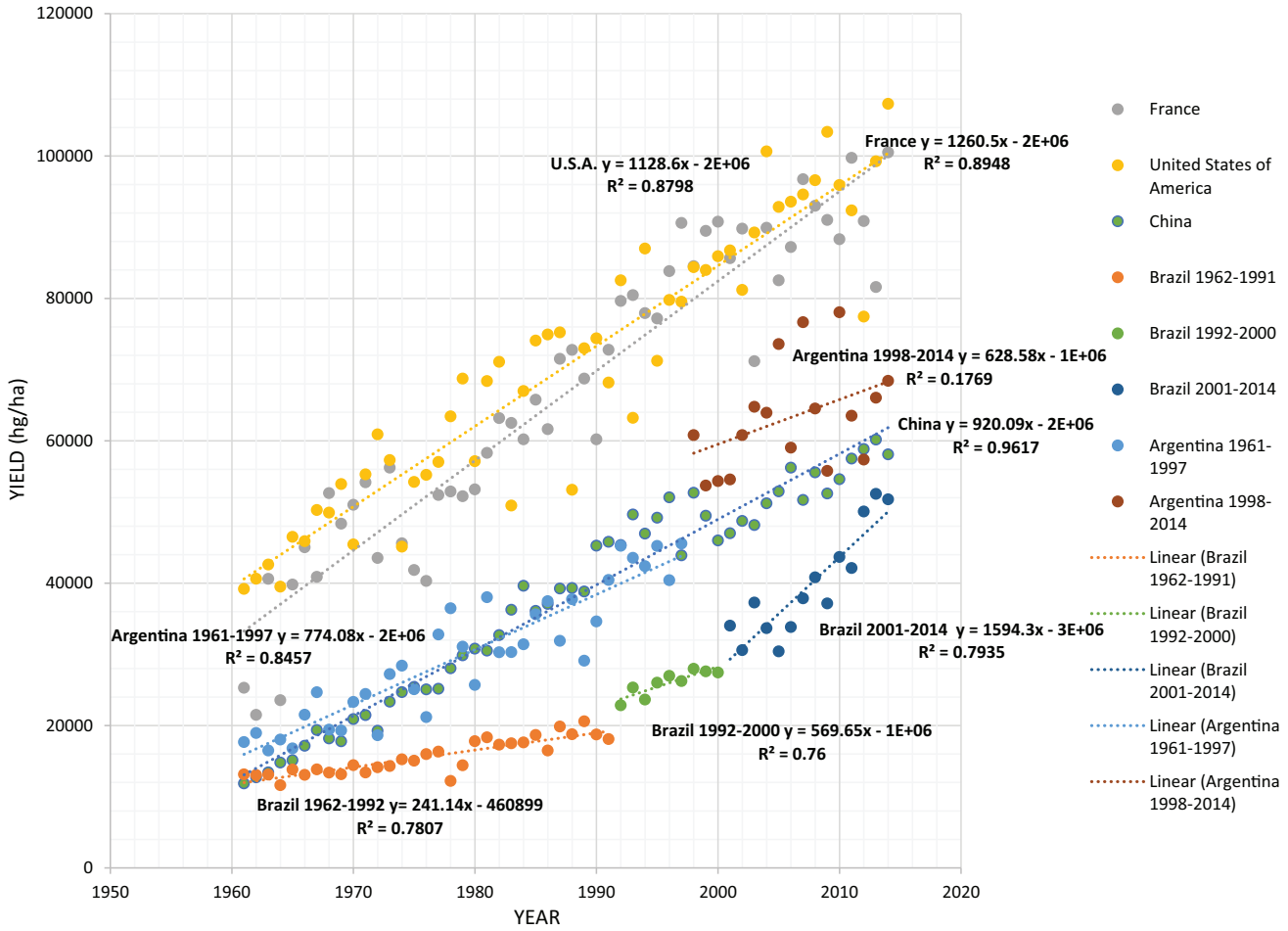

Fig. 1 Trends in maize yields 1961–2014 for the five top producing maize countries globally. Data from FAOSTAT

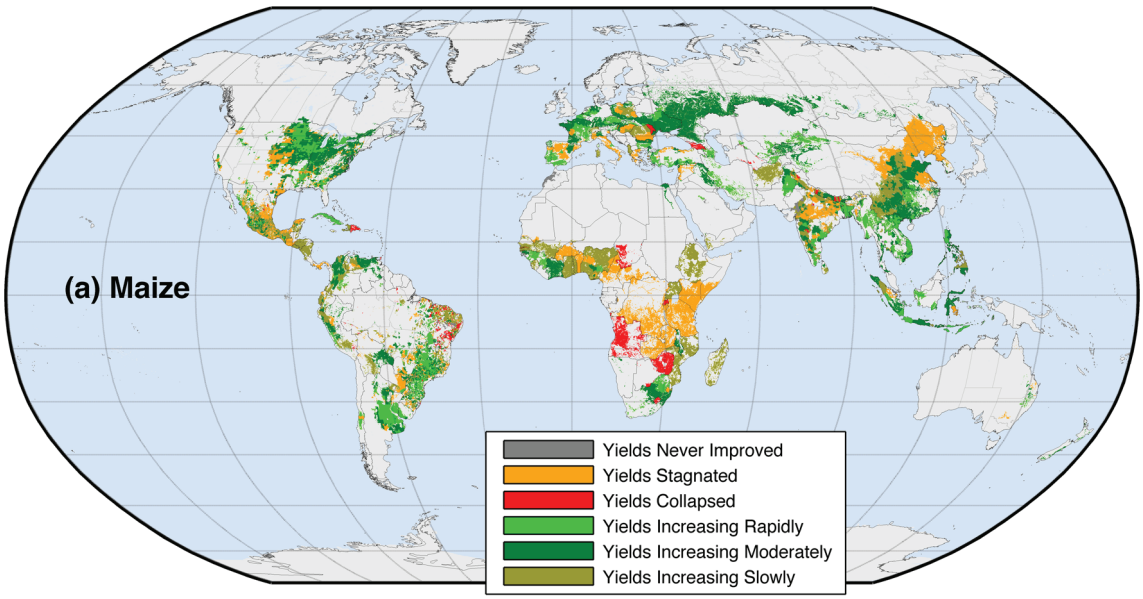

Fig. 2 Changes in maize yields for individual countries and regions on a global basis 1961–2008 Fig. 2a from Ray et al. (2012)

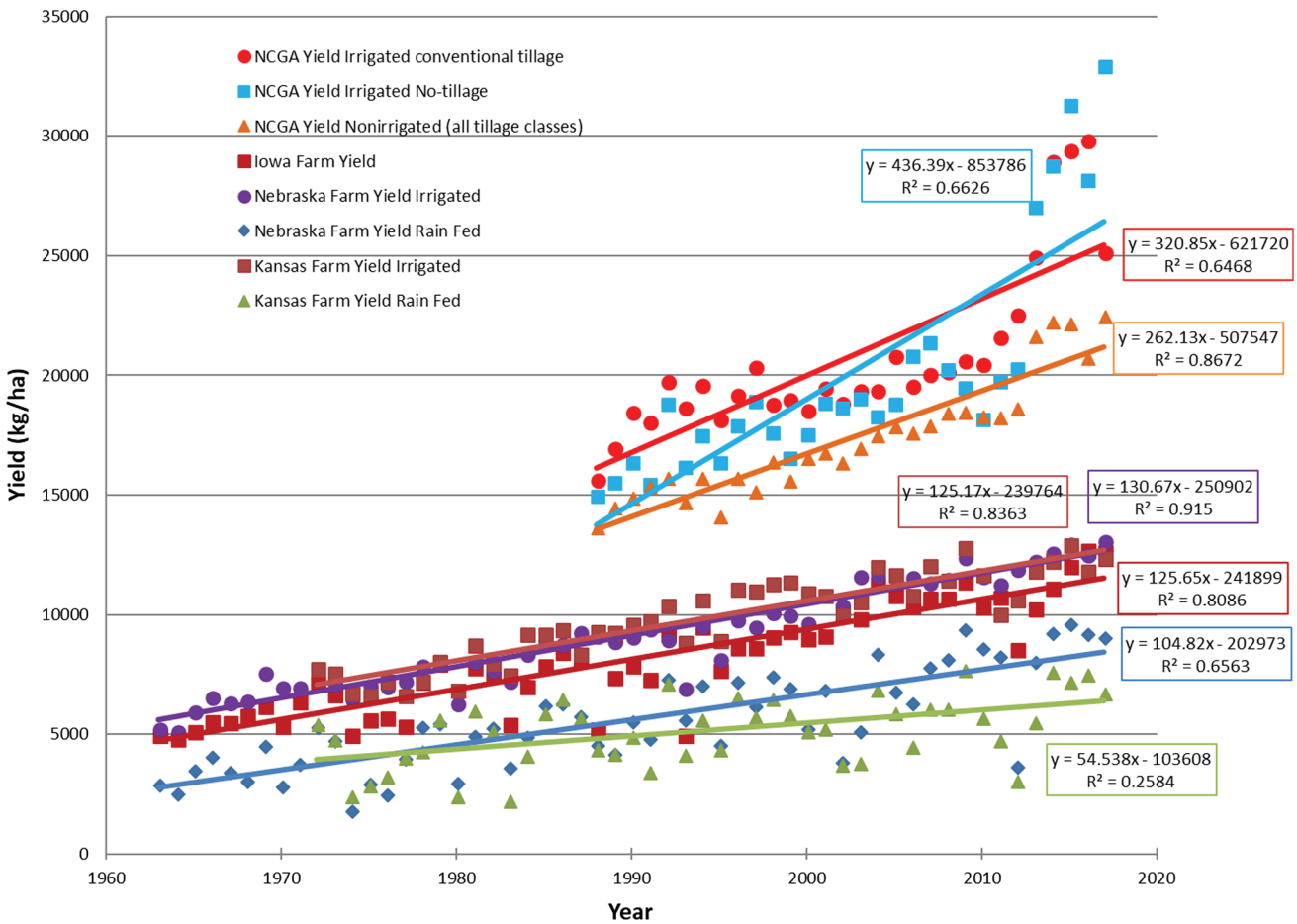

Fig. 3 Maize yields on farms during 1963–2017 in Iowa (rain-fed), Nebraska (rain-fed and irrigated), during 1947–2017 in Kansas (rain-fed and irrigated) and National Corn Grower Winners (mean of top 3 per class) reported during 1988–217 (irrigate conventional tillage, irrigated no-tillage and = rain-fed, all tillage classes)

released during 1930s–2016, indicates a single inflection point during the 1960s when single-cross hybrids replaced three- and four-way hybrids. The rate of genetic gain during 1930–1960s was 55.5 kg/ha/year, which then rose to 99.3 kg/ha/year in subsequent decades. In addition to advancing yield, maize breeders have added genetic contributions to defense against insect pests and enhanced environmental quality by facilitating conservation tillage. The rate of genetic gain *per se* may have been temporarily reduced, while selection for defensive traits increased for factors including improved germination and seedling growth in cold, wet and disease-infested soils, agro-ecological conditions that are associated with conservation tillage. Other trait deficits arise as particular hybrids gain market share, e.g., root lodging, brittle snap, disease, insect pests.

US corn grower contest winning yields have increased. However, harvests of > 25–30 t/ha are from plots located at lower latitudes, with longer growing seasons, more solar radiation and microclimates that reduce heat stress compared to the central Corn Belt. In contrast, annual contest winning yields obtained on irrigated plots in Nebraska (mean top three, each year 2015–2017) are 19.3 t/ha, similar to contest winning yields already achieved since the early 1980s (Duvick and Cassman 1999). Consequently, there is no firm evidence to show that yield potential (Yp) has increased due to changes in physiology underpinned by genetic change. The primary means by which maize yields have increased is through increased abiotic and biotic stress tolerances which avoid barrenness while allowing increased planting densities. Newer hybrids outperform older hybrids regardless of weather, drought or low nutrient stress. Nonetheless, physiological studies show that increased grain production per plant and the ability to yield under higher planting densities are independent. Consequently, advances in both attributes might provide successful breeding strategies (Gonzalez et al. 2018). Even though most selection has been practiced under fertile conditions, Haegele et al. (2013) found a relatively high rate of genetic gain under low-*N* conditions. Most genetic gain in high *N* environments could be explained by improvements in grain yield under low *N*. Mastrodomenico et al. (2018) found large genetic variation for most *N*-use traits among maize inbred lines with expired plant variety protection (PVP). Progress in NUE through breeding appears thus feasible (Haegele et al. 2013; Mastrodomenico et al. 2018).

Forecasting future climates and yield responses are notoriously complex and speculative. Heat, precipitation and solar radiation affect yield trends over years and affect annual deviations due to extreme events from year to year. Amelioration of weather effects through breeding is feasible given the repertoire of available genetic diversity for maize for climatic trends, while breeding for annual stability is more challenging, if not impossible to remedy either through genetics or agronomic practices. Enhanced levels of CO_2 do not stimulate photosynthetic accumulation of biomass due to the C4 physiology of maize, although higher levels of CO_2 can mitigate negative effects of drought stress (Leakey et al. 2006; Markelz et al. 2011). The primary effects of climate change will likely be driven by changes in heat and precipitation with various models predicting both increase and decrease in maize yield. Taken together, future breeding efforts will continue to focus on exploiting current Yp, which plateaued in yield contests of past decades, rather than increasing the Yp itself.

The objectives of this manuscript are to review important areas that have contributed to progress in maize breeding and will likely determine its future success. Specifically, we will report on progress in (1) genome analysis, (ii) germplasm diversity characterization and utilization, (3) manipulation of genetic diversity by transformation and genome editing, (4) inbred line development and hybrid seed production, (5) understanding and prediction of hybrid performance, (6) novel breeding methodologies and (7) synthesis of opportunities and challenges for future maize breeding.

Maize pan-genome: evolution of genomic resources facilitating gene identification and genetic mapping

Maize is the most widely grown agricultural crop in the world and a pre-eminent experimental model plant. This dual importance of maize is largely due to its complex and diverse genome, which has allowed researchers to better understand genetics, cytogenetics and genomics, and has offered a rich pool of genetic diversity to help breeders create improved germplasm. Maize has a medium-sized genome among the grasses with approximately 2.4 billion base pairs. There are between 30,000 and 40,000 genes in maize, with a large proportion being syntenically conserved with related grasses.

Maize maps Early maps (cytogenetic maps) (Birchler 1980) were based on the cytogenetic position of a gene relative to its location on a stained chromosome. Beginning in the late 1980s, the first "molecular" maps (Burr et al. 1988; Helentjaris et al. 1986; Weber and Helentjaris 1989; Gardiner et al. 1993) were developed using small DNA fragments (probes) that detected restriction fragment length polymorphisms (RFLPs). This early DNA mapping technique improved the coverage of genetic maps to a few hundred markers across the genome. Additional molecular marker approaches to create maps included simple sequence repeat (SSR) markers, random amplified polymorphic DNA (RAPD) markers (Lanza et al. 1997) and amplified fragment length polymorphism (AFLP) markers (Castiglioni et al. 1999). The

evolution of DNA markers as important breeding tools from RFLPs to SSRs/AFLPs and now to single-nucleotide polymorphisms (SNPs) has enabled high-resolution mapping at much lower costs. Composite genetic maps merged data from distinct mapping populations to create high-resolution linkage maps (Beavis and Grant 1991).

Progress in sequencing technologies Sanger sequencing (Sanger et al. 1977) was used for most of the early sequencing projects. Sanger sequencing produces reads of up to one kilobase (kb) in length. To produce longer stretches of sequence, a "shotgun" approach was used, where overlapping DNA was cloned and sequenced and then assembled to create contigs (i.e., contiguous sequences). These contigs were assembled to scaffolds, the framework for extended sequences. As sequencing technology steadily improved, more complex DNA entities were sequenced: genes, mitochondria, chloroplasts, single chromosomes, viruses, etc. (Heather and Chain 2016). The first plant genome (Arabidopsis) was sequenced in 2000 (Initiative 2000) and the first crop (rice) in 2002 (Goff et al. 2002; Yu et al. 2002) (Fig. 4). Prior to the first maize genome sequence, bacterial artificial chromosomes (BACs) were used to understand the structural organization of maize and to create a physical map with anchor markers on genetic maps (Yim et al. 2002). Other popular sequence data in the maize community include expressed sequence tags (EST), genome survey sequences (GSS) and random clone inserts (Dong et al. 2003). This groundwork eventually led to the maize genotype B73 being sequenced (Schnable et al. 2009), based on Sanger sequencing using the shotgun approach. B73 was chosen because it is a key founder line for various elite inbreds used in public and private hybrid cultivars.

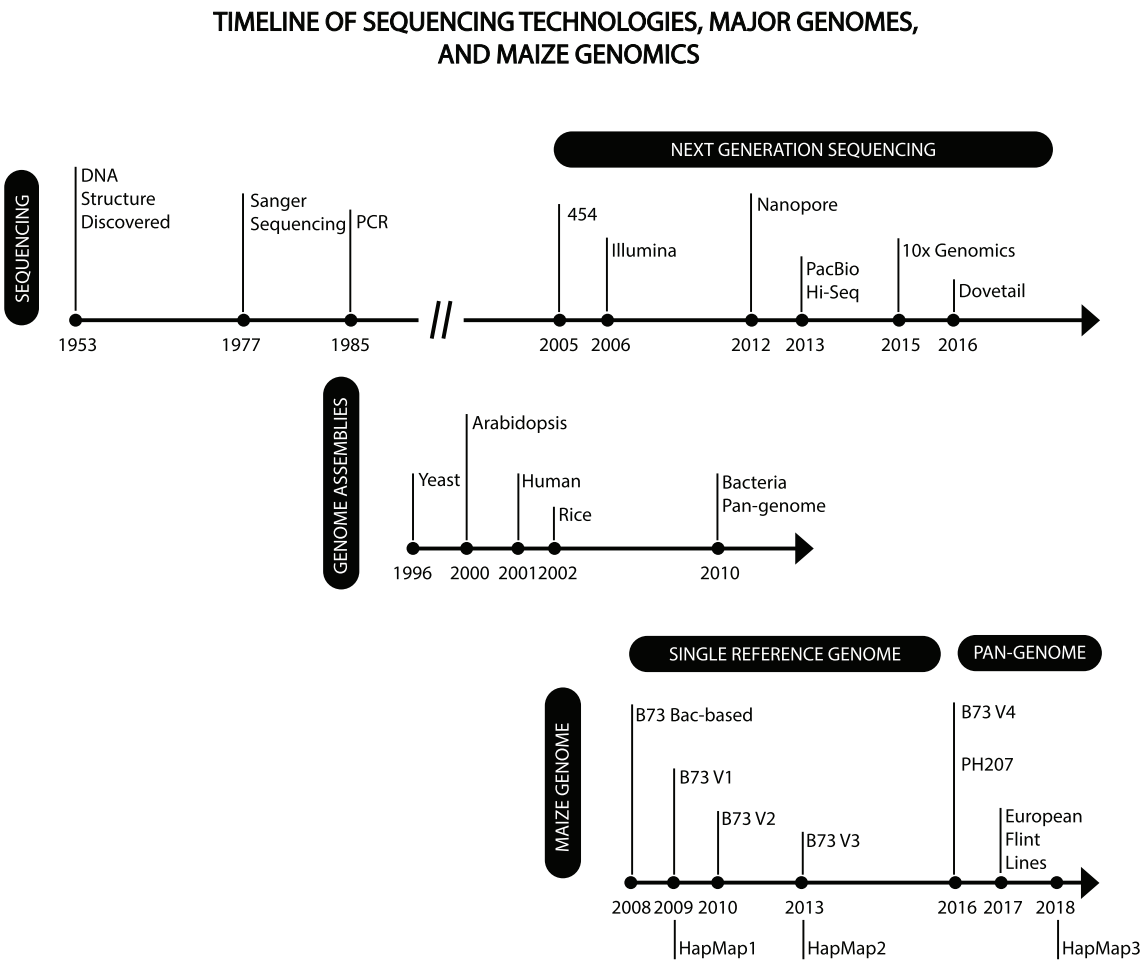

Fig. 4 Timeline of sequencing technologies, major genomes and maize genomics. The figure shows three timelines. The first timeline list the release dates of major sequencing technologies focusing on the early first-generation technologies (1950–1990) and the next-generation sequencing technologies (2000s). The second timeline shows the release dates of four major genomes (yeast, arabidopsis, human and rice) and the first reported pan-genome (bacteria). The third timeline shows the release dates of maize genomes and major genotype datasets

The second generation of sequencing technologies, also known as next-generation sequencing (NGS), was led by sequencers that were commercialized by 454 Life Sciences (Ronaghi et al. 1998) and Illumina (Voelkerding et al. 2009). These sequencers allowed for parallelized DNA amplification, which greatly increased the amount of DNA sequenced per run cycle. Current Illumina sequencers can sequence between 50- and 500-bp reads, and can produce up to 100 Gb of total sequence per lane. Many of the recent maize genome projects (Hirsch et al. 2014; Unterseer et al. 2017; Yang et al. 2017a, b) are primarily based on Illumina sequences. A third generation of sequencing technologies is emerging, such as the single-molecule real-time (SMRT) platform from Pacific Biosciences (PacBio) (Eid et al. 2009). PacBio reads can be in the 10,000 s of bp and can even reach over 100,000 bp. The major disadvantage (besides cost) is that PacBio sequences have higher error rates. Therefore, they are generally used in combination with Illumina data in bioinformatic assembly approaches. Other emerging third-generation technologies (with leading commercial providers in parentheses) include: nanopore sequencing (Clarke et al. 2009) (Oxford Nanopore Technologies), assembly of synthetic long reads (Eisenstein 2015; Li et al. 2015) (Illumina and 10× Genomics), high-throughput optical mapping (Schwartz et al. 1993) (BioNano Genomics) and chromosome conformation capture sequencing (Belton et al. 2012; Putnam et al. 2016) (Dovetail Genomics). Each of these technologies is contributing to higher-quality, more contiguous genome sequence assemblies at lower costs. The original full-genome sequencing projects took years to complete with costs measuring as high as $3 billion (human genome project). Current sequencing efforts are measured in weeks and in thousands of dollars.

B73 Maize Sequencing Project, and subsequent versions and challenges Maize is a difficult genome to sequence and assemble due to its complex and repetitive composition. Maize has hundreds of thousands of long terminal repeats accounting for about 85% of the genome (Huang et al. 2012). In addition, maize is a paleopolyploid. One genome duplication event occurred around 5–12 million years ago, and another one predates the divergence of cereal crops around 70 million years ago (Schnable and Freeling 2011; Woodhouse et al. 2010). The first maize genome sequence (Schnable et al. 2009) was completed by the Maize Genome Sequencing Consortium (MGSC). This genome assembly (B73 RefGen_v1) contained 2048 Mb in 125,325 sequence contigs (N50 of 40 kb), forming 61,161 scaffolds (N50 of 76 kb), and was anchored to a high-resolution genetic map (Wei et al. 2009). The structural annotation included a total of 32,540 high-confidence protein-coding genes. Since then, the B73 sequence has undergone three major updates (RefGen_v2, RefGen_v3, RefGen_v4). The latest B73 RefGen_v4 (Jiao et al. 2017) was based on PacBio sequencing and high-resolution optical mapping and is the most accurate assembly of maize to date.

For the past 8 years, B73 has been the only publicly available sequenced maize genome and has been the focus and the representative *Z. mays* reference genome at the public information resources MaizeGDB (Andorf et al. 2016), Gramene (Tello-Ruiz et al. 2018), Ensembl Plants (Bolser et al. 2017), GenBank (Benson et al. 2013), EMBL-EBI (Chojnacki et al. 2017) and Phytozome (Goodstein et al. 2012). The structural annotations for B73 are used in most maize experiments and are cited in the majority of maize publications.

HapMap and Diversity projects To examine the genetic diversity of maize, thousands of maize lines have been genotyped and aligned to B73. SNPs are usually determined relative to B73 (Jiao et al. 2012; Ganal et al. 2011; Lu et al. 2015; Romay et al. 2013; Wallace et al. 2014). The most comprehensive dataset generated began in 2009 as a complementary study to the B73 reference genome. Millions of sequence variations across 27 maize genome lines were identified to create a first-generation haplotype map of maize (HapMap) (Gore et al. 2009). The study found areas of suppressed recombination near centromeres and hundreds of regions associated with geographic adaptation. HapMaps are especially important in maize due to the high variation across any two maize lines, including extensive presence/absence variation at the gene level between inbred lines. Subsequent versions, HapMap2 (Chia et al. 2012) and HapMap3 (Bukowski et al. 2018), expanded the number of lines and increased resolution by additional SNPs. HapMap2 found 55 million SNPs among 103 maize and teosinte (*i.e.*, wild maize) lines. HapMap3 identified 83 million variant sites for 1218 maize lines.

Maize pan-genome In addition to B73 RefGen_v4, complete reference-quality maize genomes released in the past year include W22 (Springer et al. 2018), PH207 (Hirsch et al. 2014), CML247 (Lu et al. 2015), Mo17 (Sun et al. 2018), *Z. mays* Mexicana (Yang et al. 2017a, b) and the European Flint lines EP1 and F7 (Unterseer et al. 2017). Soon several more maize and *Zea* lines are expected to be released, including B104 (USDA-ARS/Iowa State University), Ki11 and NC350 (Doreen Ware, USDA-ARS) and a *Zea diploperennis* genome (Matthew Hufford, Iowa State University). At this rate, dozens, if not hundreds, of sequenced maize genomes will be available in the near future.

These newly assembled genomes are integrated into a maize pan-genome, a concept first illustrated in bacteria (Donati et al. 2010; Liu et al. 2014; Tettelin et al. 2005). Pan-genomes contain two types of gene models: core genes (or accessory genes) (Segerman 2012) and pan-genes or

dispensable genes (Li et al. 2014; Vernikos et al. 2015). For maize, core genes are found in all lines of maize, whereas pan-genes are found at least once but not in all maize lines. An initial maize pan-genome study (Hirsch et al. 2014) found that approximately 38% of all annotated B73 reference genes were present in all 503 of the maize inbred lines examined. This percentage is low in comparison with the percentage of core genes found in 3000 cultivars in one rice study (47%) (Sun et al. 2017) and to the percentage of core genes (80%) found in a soybean pan-genome made up of seven cultivars (Li et al. 2014). However, as the number of core genes and pan-genes are relative to the number of cultivars studied, we predict the expected number of pan and core genes in maize will change as more maize accessions are sequenced and compared.

When constructing a pan-genome, accuracy depends on assembly and annotation quality, orthologue detection methods and diversity of the selected lines (Golicz et al. 2016). Visual representation of a pan-genome is still an open question with many challenges (Consortium 2018; Golicz et al. 2016). A large and complex genome such as maize presents more challenges compared to small genomes. Diversity involves differences not only in sequence but also in the position of orthologues within the pan-genome, as well as in copy number. Certain gene families are positionally dynamic and tend to reside in tandem arrays. The simplest approach to create a maize pan-genome is to select a reference genome against which all other genomes will be compared. This is the approach taken with initial maize pan-genomes, though this approach limits the study of full maize diversity, since it excludes genes that might be present in other maize lines but not in the reference genome. An ideal maize pan-genome would include positional information for all contributing maize lines, representing the total positional diversity in maize. De Bruijn graphical representations (Paten et al. 2017), such as a practical haplotype graph (PHG), may offer solutions to this problem.

Gene and quantitative trait locus (QTL) mapping One of the driving forces to establish genetic maps, sequence genomes, etc. is the interest in the identification of genes or QTL affecting traits of interest. While genetic mapping has been used for more than 100 years, a limitation has been the low number of genetic markers, until the advent of molecular markers in the 1980s. In biparental populations using few hundred markers covering the genome, genes, mutant phenotypes and QTL were mapped based on their linkage to RFLP markers. Genetic fine mapping along with physical maps enabled map-based gene isolation in the 1990s. Progress in molecular techniques was accompanied by progress in the development of statistical methods. For QTL mapping, the initial simple and interval mapping has been replaced by composite and multiple interval mapping approaches (Jeffrey and Lübberstedt 2014). A limitation of biparental populations (e.g., F_2-derived) of low genetic resolution was initially addressed by repeated intercrossing before the development of mapping families. The intermated B73 and Mo17 (IBM) population (Lee et al. 2002) was established as community resource, with more than 2000 markers and representing two of the major heterotic groups used in US maize germplasm (Coe et al. 2002; Cone et al. 2002; Lee et al. 2002). Subsequently more sophisticated multiparental mapping populations were developed, capitalizing on sequence-based markers at high density. Most prominently, the maize nested association mapping (NAM) population was created (Yu et al. 2008) by crossing B73 to 25 diverse maize lines (a.k.a. founder lines). For each of the 25 F_1s, 200 recombinant inbred lines (RILs) were developed. With a total of 5000 RILs, and combined linkage and association mapping, the maize NAM population has been a valuable public resource to map genes for various complex traits such as flowering time (Buckler et al. 2009) and disease resistance (Kump et al. 2011).

Based on concepts in human genetics, candidate gene-based and later genome-wide association studies (GWAS) have become increasingly popular, enabled by high-density markers and sequenced genotypes. The main strength of GWAS populations is their low LD and thus the ability to map loci at high resolution, potentially pinpointing causative genes. For example, the above-mentioned HapMaps were used to identify SNPs associated with agriculturally important traits including leaf architecture (Tian et al. 2011; Li et al. 2012b), resistance to Northern and Southern Leaf Blight (Kump et al. 2011; Poland et al. 2011), plant height, flowering time (Li et al. 2016; Peiffer et al. 2014) and ear rot disease resistance (Zila et al. 2014). Other GWAS panels in maize include the Ames panel (Romay et al. 2013; Pace et al. 2015) and the BGEM panel (Sanchez et al. 2018; Vanous et al. 2018). The identification of genes of interest did undergo a major paradigm shift from being an exclusively forward genetic approach to being increasingly a reverse genetic approach based on rapidly accumulating sequence and gene function information. Moreover, with an increasing number of genotyped panels and mapping populations, gene and QTL mapping efforts shifted from genotyping to phenotyping and analysis of large datasets.

The Maize Genetics and Genomics Database (MaizeGDB) integrates mapping data from a wide range of genetic maps currently hosting over 2000 genetics maps. A universal composite map (i.e., Genetic Map) is updated and maintained at MaizeGDB (Andorf et al. 2016; Lawrence et al. 2008). Table 1 shows a snapshot (August 2018 release) of the counts of major data types available at MaizeGDB. While the numbers of mapping studies and identified QTL and loci are overwhelming, the statement of Bernardo (2009) is certainly still valid: "…the vast majority of the

Table 1 Data types available at MaizeGDB. The table lists 12 major data types at MaizeGDB with the counts of each. Genomes are the most recent version of reference-quality genome assemblies. Gene models are the structural gene predictions for each genome assembly. Genes are loci in the maize genome identified as a gene. Annotated genes are genes that are associated with a phenotype or have a gene ontology term. Genetic maps are maps created by the maize community. Loci/QTL are points, probes, QTL, etc. identified in the maize genome that have functional or regulatory relevance. Markers are loci in the genome used as molecular markers. SNPs are single-nucleotide polymorphisms in the maize genome. Count information is for an allele for each position per germplasm. Germplasm are records of germplasm or genetic stocks. Gene expression is based on RNA-Seq experiments. Phenotype terms are unique descriptions for phenotypes

Features in MaizeGDB	Numbers available August 2018
Genomes	13
Gene models	598,794
Genes	10,149
Annotated genes	2068
Genetic maps	2117
Loci/QTL	214,464
Markers	771,136
SNPs	117 billion
Germplasm	66,825
Gene expression	17 studies, 158 tissues/conditions
Phenotype terms	1121

favorable alleles at these identified QTL reside in journals on library shelves rather than in cultivars that have been improved through the introgression or selection of these favorable QTL alleles." Alternative approaches to marker-assisted selection such as genomic selection (see below), not requiring mapped loci, seem to be of more practical relevance for plant breeders at this time.

Maize germplasm: diversity and utilization

Origins and taxonomic organization The origin of maize was hotly debated until the late 1970s after which genetic studies, including the use of molecular markers and comparative DNA sequence data allowed breakthroughs in the taxonomy and phylogeny of maize and its wild relatives, including the identification of specific loci involved in the domestication process. Maize was domesticated in the tropical lowlands of southwest Mexico with subsequent introgression from teosinte (Matsuoka et al. 2002; Piperno et al. 2009; van Heerwaarden et al. 2011; Hufford et al. 2013). Maize diversified under genetic drift and selection as it was carried through a diverse habitats during its spread by humans both south and north from its origin, including its arrival in the southwestern region of North America by 2260 BC (Merrill et al. 2009). The initial selection for adaptation to a temperate environment then occurred during the subsequent 2000 yrs in North America (Bouchet et al. 2013; Swarts et al. 2017). Further introductions occurred into the USA, Europe, Africa and Asia in the 16th C (Vigouroux et al. 2008; Bedoya et al. 2017; Edmeades et al. 2017). Further climatic adaptation leads to the development of European flint landraces which involved different genetic loci to those associated with temperate adaptation of dent germplasm in North America (Unterseer et al. 2016).

Characterization of germplasm provides an improved basis to inform plant breeders and conservators about genetic resource diversity. Morphological descriptions of races or "group(s) of related individuals with enough characteristics in common to permit their recognition as a group" (Anderson and Cutler 1942) were initiated in 1919, with further studies during 1943–1952 culminating in a series of "race bulletins" for Mexico, central and South America 1952–1963 (Brown and Goodman 1977). Other publications describing maize races for Europe and Asia are cited by Brown and Goodman (1977). Collectively, some 285 maize races have been described in the several "Races of maize" publications, see, for example, Wellhausen et al. (1952), although Hallauer and Miranda (1981) conjectured these collections may represent 130 distinct types. In contrast, more than 300 maize races are reported to be represented in the collection maintained at CIMMYT (www.genebanks.org/resources/crops/maize/). Six races have achieved global economic importance: Mexican Dents, Corn Belt Dents (CBD), Tusons, Caribbean Flints, Northern Flints and Flours, and the Catetos (Argentinean Flints), although several other races are important regionally (Goodman 1978). Comparisons utilizing molecular marker data mostly support previously assigned racial groupings (Liu et al. 2003: Lu et al. 2011; Mir et al. 2013). Comparisons of molecular marker or DNA sequence data allow global views of taxonomic and phylogenetic relationships of maize genetic resources (Lu et al. 2009). This capability allows genetic diversity present in one country or region to be compared in a global context, both quantitatively and particularly when linked to phenotypes, in qualitative terms also. This holds true, for example, for African maize in a global context (Westengen et al. 2012), similarly for maize utilized in China (Zhang et al. 2016) and likewise for maize cultivated in Europe (Tenaillon and Charcosset 2011; Brandenburg et al. 2017), including development of breeding strategies to further broaden the genetic base of European maize with continued introductions of US Corn Belt germplasm (Reif et al. 2010). Similarly, this is valid for the CBDs, which have achieved global usage and which comprise a relatively diverse germplasm base due to their origins from highly differentiated Northern Flints (NF) and Southern Dents (SD) (Doebley et al. 1988; Dubreuil et al. 1996; Troyer 1999, 2006; Vigouroux et al. 2008; Bauer et al. 2013; Giraud et al. 2014; Unterseer et al. 2016). Nevertheless,

the CBDs comprise a minority of germplasm diversity that is represented among maize landraces globally (Mir et al. 2013), most of which is found in tropical maize (Lu et al. 2011). Patterns of genetic diversity and phylogenies of maize in the American continent where maize was domesticated and diversified during later millennia are available to conservators, geneticists and developers of new germplasm (Vigouroux et al. 2008; Bedoya et al. 2017). Comparisons of genetic diversity between tropical, subtropical and temperate maize germplasm are available (Reif et al. 2003; Laborda et al. 2005; Lu et al. 2011). Comparisons of genetic diversity within and among breeding programs can be made (Inghelandt et al. 2010; Nelson et al. 2016). Molecular marker technology has evolved rapidly and substantially from the mid-1970s when 21 isozymic loci and several blocks of zein coding loci were available as molecular markers through to today when DNA sequence data are utilized. The culmination of these developments to the use of sequence data is very important because opportunities are now available to genetically characterize inbred lines, hybrids and populations, including landraces utilizing a common global genetic "language."

Studies of hybrid diversity It is challenging, yet vital to monitor genetic diversity during selection (NRC 1972, 1993; Rogers and McGuire 2015; Brown and Hodgkin 2015; CGC 2018). However, the ideal is difficult, if not impossible to achieve due to (1) the complexity of the maize genome, (2) G × G and G × E interactions in phenotypic expressions and (3) the difficulty in predicting useful future traits. Consequently, pedigree and molecular markers or sequence data provide useful surrogates. Temporal changes for maize diversity deployed on farms are available for France (Le Clerc et al. 2005), for Zimbabwe, Zambia and Malawi (Magorokosho 2006), and for the USA, although during the past 2–3 decades only for inbred lines after expiration of their terms of PVP and utility patent protection have expired (Nelson and Goodman 2008; Romay et al. 2013; Beckett et al. 2017). Historically, most studies of diversity have focused on US hybrid maize due to its relative longevity in cultivation and the importance of the CBDs to global maize production, not only in the USA but in many other countries. Six pedigree-based studies on the use of public inbred lines carried out between 1956 and 1986 (Darrah and Zuber 1986), showing change in diversity in time, although by 1984 only three publicly bred inbreds contributed >1% to hybrid seed production. These surveys seemed to indicate that diversity of the US maize crop might be narrowing (NRC 1972; Zuber and Darrah 1981). However, more positive responses were received about available diversity (Duvick 1984), reflecting the understanding that diversity resides in breeding programs as a whole rather than just the commercial portfolio at any single place and time (Duvick 1984).

The optimal way to meaningfully monitor diversity is to apply molecular markers directly to hybrids coupled with information on their relative usage. Molecular marker surveys of widely used hybrids during the 1986 and 1990 harvests (Smith and Smith 1991; Smith et al. 1992) showed that many hybrids (46–48%) grouped with others, including open pedigreed hybrids at > 95% thresholds of similarity. Nonetheless, 5–6 out of 18 companies had > 50% of their widely grown hybrids that could be categorized as "unique." For the 1990 harvest (Smith et al. 1992), US hybrids were associated into two large clusters: The first comprised 91% of the hybrids developed by Pioneer Hi-Bred International (now Corteva Agriscience), two were > 95% similar, three hybrids developed by DeKalb, and one each by Funk Seeds and Asgrow (now all part of Bayer Crop Science). The second comprised all other hybrids including seven subgroups of hybrids.

Status of genetic diversity in US maize breeding and agriculture today Although legal restrictions have stymied assessments of temporal trends in cultivated maize diversity, some insights can still be gleaned from published data. Mikel and Dudley (2006) collated pedigrees of proprietary US inbred lines from information in the PVP and patent databases. Comparisons of a series of progenitor lines indicate that major contributions are from public breeding programs and by Pioneer Hi-Bred International (PHI), including via Holden Foundation Seeds (now Bayer Crop Science). When additional inbreds are added, then stiff stalk contributions from DeKalb and one inbred developed by Northrup King (now Syngenta) occur. These data beg the question: What happened to diversity previously developed by at least five other breeding programs as exemplified by the molecular marker-based assessments of uniqueness of several hybrids widely used during the 1986–1990 time frame (Smith and Smith 1991; Smith et al. 1992)? Some diversity seems to have been lost, and thus, major breeding programs are becoming genetically more similar. There is continued heavy usage of B73, PH207 and PH595 descendents, sourced, not only via PHI hybrids 3180, 3535, 3737, but also directly via proprietary inbred lines (Garing 2000; Larkins 2000). Continued development of US hybrids has increasingly apportioned genetic diversity between heterotic groups (Feng et al. 2006). In contrast, van Heerwaarden et al. (2012) demonstrated a narrowing of SNP haplotype diversity within each of three US heterotic pools.

Genetic diversity in US maize: present situation and future Given evidence of an overall narrowing of genetic diversity in US maize germplasm, the more effective use of a broader genetic resource base is an important strategy to pursue.

Greater usage of CBDs globally is one example, but risks of further global dependence upon a relatively limited genetic base should be hedged by utilizing programs that can more effectively characterize hitherto underutilized, including exotic genetic resources. However, there are very few examples where exotic germplasm has been successfully incorporated into the CBDs with marginal use of temperate exotic germplasm (usually 2–6%, occasionally 12–25%) and minimal use of tropical germplasm (0.1–5%) in a few hybrids (Goodman 1999, 2005). The Maize Crop Germplasm Committee (MCGC) expressed concerns about the vulnerability of genetic diversity in US maize stating that "the genetic health of the maize crop is a matter of National security" (MCGC 2016). Chief recommendations included: (1) international collaborations to screen for resistance to diseases not yet found in the USA, (2) genetic diversity of the U.S. maize crop should be evaluated using DNA-based tools, (3) regeneration and characterization must be increased, (4) additional collections of landraces, populations, wild relatives and inbred lines from programs are needed before closure and (5) expansion of germplasm enhancement. A reduction in useful genetic diversity will ultimately result in a decline in the rate of genetic improvement unless remedial measures are taken. The rate of decline accelerates as inbred development becomes more effective (Gaynor et al. 2017). Programs to increase diversity require "long-term commitment and appropriate breeding strategies, and may be assisted by DNA marker technologies" (Holland 2004). Programs designed specifically to adapt and characterize exotic germplasm for the purpose of identifying new useful diversity are termed "pre-breeding." The international scope of breeding programs provides prospects whereby breeding in one location is "pre-breeding" for other global locations. For example, introgression of temperate germplasm into tropical hybrids developed in Brazil may, in turn, help identify useful tropical germplasm for use in temperate locations. An exotic germplasm introduction program initiated at NCSU takes advantage of cycles of inbreeding to develop tropical inbred lines as a means of reducing alleles that may have deleterious effects. Further cycles of pre-breeding then occur in a temperate climate (Goodman 2005; Nelson and Goodman 2008; Gardner 2012; Nelson et al. 2016). Readers are recommended to Smith et al. (2017) for further information on the importance and challenges of utilizing wild and other exotic genetic diversity in maize breeding, case studies demonstrating the use of exotic genetic resources and critical issues faced by genebank curators now and in the immediate future. New breeding strategies including the use of molecular markers and precision phenotyping provide improved opportunities to more effectively access diversity (Yu et al. 2016) and see Seeds of Discovery (SeeD) (Prasanna 2012). Breeding programs that are primarily aimed at developing improved hybrids can also be modified to reverse the decline in genetic variance (Gaynor et al. 2017; Voss-Fels and Snowdon 2016).

Manipulation of genetic diversity: mutagenesis, transformation, genome editing

Induced mutagenesis Manipulation of genetic diversity in maize can be achieved through various approaches such as hybridization with sexually compatible wild relatives (Mangelsdorf 1961), treatment with physical and chemical mutagens (Bird and Neuffer 1987), transposable elements (May et al. 2003; Neuffer et al. 2009), transgenesis (Wang et al. 2003a, b) and genome editing (Gao 2018) (Table 2). All these approaches can cause gene mutations and rearrangements. Besides its utility in broadening genetic variation, mutagenesis is also a useful tool to understanding gene function. For instance, X-rays were used to induce mutations at the yellow–green (*Yg*) locus (Dollinger 1954), and ethyl methanesulfonate (EMS) was used to determine the function of the Sh_1 protein involved in endosperm development (Chourey and Schwartz 1971).

Attempts to introduce physical changes to the maize genome using irradiation started as early as the 1930s (Stadler and Sprague 1936). This work helped demonstrate that shorter wavelengths (~2600 Å) of non-ionizing radiation such as UV are more effective in inducing DNA damage (Stadler and Sprague 1936). In addition, toxic compounds such as mustard gas were evaluated for their effectiveness as mutagens in maize (Gibson et al. 1950). However, more success was attained with UV radiation (Stadler and Uber 1942) and ionizing radiation, such as X-rays and gamma rays (Neuffer 1957; Sarvella and Grogan 1967). In this early work (Gibson et al. 1950; Sarvella and Grogan 1967; Stadler and Uber 1942), there were limitations with respect to exposing target tissue and cells to the mutagen to effectively generate heritable mutations. For example, maize pollen was the main target for the application of UV radiation and mustard gas likely due to the abundance and ability to pass the mutations to the progeny.

Induction of mutations in maize with chemical mutagens, for example, EMS and various *N*-nitroso compounds has also been carried out for a while (Amano and Smith 1965; Bird and Neuffer 1987; Williams 2016). EMS is more effective than non-ionizing and ionizing radiation (Neuffer and Fiscor 1963; Neuffer et al. 2009). When EMS is used in combination with carriers such as paraffin (Neuffer and Coe 1978) or mineral oil (Neuffer 1994), less damage to the germ cells is experienced, resulting in increased mutation frequency (Brunelle et al. 2017). TILLING (Targeting Induced Local Lesions IN Genomes) uses chemical mutagenesis methods to create libraries of mutagenized seed that are later screened using high-throughput approaches for the discovery of useful mutations. With an increase understanding of

Table 2 Progress in the use of mutagenesis, transgenesis and genome editing to broaden genetic diversity in maize

	1930–1950	1950–1970	1970–1990	1990–2010	2010–2030
Induced mutagenesis	Irradiation of maize pollen with shorter wavelengths of UV light induces genetic mutations Use of mustard gas to induce mutations in maize Mutations could be induced by transposable elements	Ionizing radiation, such as X-rays and gamma rays can induce chromosomal changes Hybridization of maize with teosinte has a mutagenic effect Treatment of maize pollen with ethyl methanesulfonate (EMS) is more effective than X-ray and UV irradiation	Use of paraffin oil improves delivery of EMS to pollen grains and reduces damage to germ cells Use of the Robertson's mutator (Mu) system for production of new maize mutants Tissue culture induces genomic rearrangements in maize	Transposon tagging for cloning and functional characterization of maize genes TILLING	Treatment of maize seeds with gamma irradiation, *N*-ethyl-*N*-nitrosourea and sodium azide expands the possible mutation spectrum EMS-induced embryo-specific mutation frequency is three times higher
Transgenesis		First attempt to transform maize by direct DNA injection	Stable transformation of maize callus via electroporation of protoplasts Discovery that Agrobacterium can transfer DNA to maize Transgenic maize plants with resistance to antibiotic regenerated from protoplasts. However, the plants were sterile	Production of fertile transgenic maize with herbicide and insect resistance High efficiency transformation of maize by Agrobacterium Polyethylene glycol (PEG)-mediated protoplast transformation	
Genome editing				Specific mutagenesis of maize *IPK1* gene with ZFN technology	First application of CRISPR/Cas technology in Use of Cas9 and guide RNA for gene editing in maize Targeted genome modification by CRISPR/Cas9 Use of TALENs to induce heritable and site-specific mutations in maize Establishment of cost-effective CRISPR platform for targeted mutagenesis in maize

sequence—function relationships—valuable alleles might be identified in respective TILLING populations in a more targeted way by reverse genetic approaches. Till et al. (2004) obtained 17 independent EMS-induced mutations from a population of 750 maize plants derived from mutagenized pollen. Such novel mutations are useful in dissecting complex traits such as seed number (Bommert et al. 2013).

Transposable elements or transposons are DNA sequences capable of migrating around the genome and may induce various chromosomal mutations and genetic variation. The early works of Rollins Emerson (1917), Barbara McClintock (1950), Marcus Rhoades (1938) and Peter Peterson (1953) on transposable elements paved the way for their widespread use in maize research. Recent discoveries suggest a role for transposable elements in maize gene regulation in response to stress conditions (Makarevitch et al. 2015). Examples of transposable element systems in maize include activator–dissociation (*Ac-Ds*), suppressor–mutator (*Spm*) and Robertson's mutator (*Mu*) (Robertson 1957). Although the use *Ac-Ds*, *Spm* and *Mu* transposon systems became useful methods to study gene function in maize (Neuffer et al. 2009), the loss of transposon activity can lead to the suppression of large numbers of mutations in the F_2 population, making it difficult to understand the function genes of interest (May et al. 2003). In recent years, transposon-based genetic resources such as mapped *Ac/Ds* families and UniformMu have been established and are available through MaizeGDB (https://www.maizegdb.org/) and Maize Genetics Cooperation Stock Center (http://maizecoop.cropsci.uiuc.edu/). UniformMu (McCarty et al. 2013) is a population of mutator-induced mutants in a highly uniform W22 background. Over 95,871 unique germinal insertions were mapped in over 14,000 seed stocks. Insertion positions are available for both B73 and W22. The Ac/Ds (Vollbrecht et al. 2010) resource is a collection of sequence-tagged Ds insertions in W22-derived inbred lines generated by aligning 2072 Ds flanking sequences against B73. Both resources display the insertions as genome browser tracks at MaizeGDB with tools to order the seed stock containing the insertion.

Transgenesis Despite the important value of induced mutations, the process of mutation breeding is labor-intensive and time-consuming. Some of the breeding objectives can be achieved through in vitro tissue culture processes and genetic transformation (Lee and Phillips 1987). First attempts to transform maize involved direct injection of DNA into tissue, but without success (Coe and Sarkar 1966). Twenty years later, progress in biotechnology resulted in stable transformation of maize (Fromm et al. 1986) and the discovery that Agrobacterium enabled transfer of DNA to maize cells (Grimsley et al. 1987). The first transgenic maize was developed by protoplast transformation (Rhodes et al. 1988), albeit infertile. The first fertile transgenic maize was developed by particle bombardment of embryogenic suspension culture (Gordon-Kamm et al. 1990) and protoplast transformation (Golovkin et al. 1993). Since then, transgenic maize plants primarily carrying insect resistance and herbicide tolerance traits have been commercialized in many countries and are one of the major successes of this technology in the last century. In 2016, transgenic maize occupied 59.7 mio. hectares globally (ISAAA 2017).

In the mid-1990s to early 2000s, more robust protocols were developed for *Agrobacterium*-mediated maize transformation (Ishida et al. 1996) and polyethylene glycol (PEG)-mediated protoplast transformation (Wang et al. 2000). Maize transformation protocols are highly genotype dependent. Therefore, the identification of transformable genotypes was a primary challenge. High type II callus production (Hi II) became one of the most widely used families for maize transformation because of its ability to produce highly transformable calluses (Zhao et al. 2001; Frame et al. 2002). However, Hi II is a segregating family, which complicates gene function studies. For that reason, the discovery that overexpression of maize genes encoding baby boom and Wuschel morphogenic regulators can enable leaf cell transformation of recalcitrant genotypes (Lowe et al. 2016) is an important milestone for maize transformation.

Genome editing Site-directed nuclease systems have been developed for targeted genome editing for more than two decades. The application of genome editing technologies is expected to generate new genetic variation in maize for both basic research and development of improved commercial germplasm. Current genome editing tools use nucleases to induce DNA double-strand breaks (DSBs). These tools include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced palindromic repeats (CRISPR)-CRISPR-associated (Cas)-CRISPR-Cas systems (Georges and Ray 2017).

The application of ZFN technology resulted in reduction in seed phytate content via specific targeting of one of the inositol phosphate kinase (IPK) homologues (Shukla et al. 2009). However, a challenge with the ZFN technology is its limited ability to generate a high frequency of mutations, making it difficult to identify the mutated alleles (Puchta and Hohn 2010).

TALENs are similar to ZFNs and comprise a non-specific *Fok*1 nuclease fused to a DNA-binding domain. However, the biggest challenge with the use of TALENs is engineering a highly specific TALE domain to avoid off-target DNA cleavage. Such non-specific DNA editing may have deleterious results making it difficult to obtain a desirable mutation. Nonetheless, using a combination of gene promoters, heritable and site-specific DNA changes in the maize *glossy2* (*gl2*) locus were generated by the TALEN approach using

Agrobacterium transformation of the B104 inbred (Char et al. 2015).

In maize, CRISPR technology has been applied to modify various traits such as male sterility, lignin biosynthesis, herbicide tolerance, secondary metabolism, grain composition and drought tolerance (Chilcoat et al. 2017). The first use of Cas9/gRNA for genome editing in maize targeted the maize IPK gene (*ZmIPK*) using PEG-mediated protoplast transformation (Liang et al. 2014). Similar experiments involved transformation of immature embryos of the B73 inbred line (Xing et al. 2014). In addition, CRISPR/Cas9 has been used to induce mutations and replace or add genes in maize using biolistic maize transformation (Svitashev et al. 2015). Char et al. (2017) showed that co-infection of two Agrobacterium strains harboring distinct Cas9/gRNA can generate transgenic plants with mutation rates as high as 70%. Even though the CRISPR-Cas technology enabled modification of various traits in maize (Svitashev et al. 2015, 2016; Shi et al. 2017; Char et al. 2017), the native maize GOS2 promoter was used by Shi et al. (2017) to both replace and supplement the native ARGOS8 promoter. Variants with altered expression of ARGOS8, a negative regulator of ethylene responses, showed yield gains under drought stress, with no yield penalty under well-watered conditions (Shi et al. 2017).

Future directions Maize harbors a vast amount of spontaneous mutations that can be leveraged to create adapted varieties and inbred lines (Bird and Neuffer 1987). However, this genetic variation may not be available in elite germplasm. Genome editing methods are precise and faster to attain desirable genome changes without lengthy backcross procedures. Coupled with genotype-independent maize transformation, genome editing technologies could become a prevalent approach to introduce specific genetic changes for organizations with negotiated access to genome editing technologies and financial resources to address regulatory requirements.

Inbred line and hybrid seed production

Maize has a convenient reproductive organization with separate male and female flowers on the same plant enabling both inexpensive self-pollination for inbred line development and controlled cross-pollination for hybrid seed production. This is likely one of the reasons why hybrid breeding was first implemented in maize after discovery of heterosis. Efficient hybrid breeding requires methods that (a) quickly generate homozygous and homogeneous lines and (b) enable cost-efficient seed production. Inbred lines developed by continuous self-pollination have been largely replaced by doubled haploid (DH) lines. Hybrid seed production was initially done by individual hand crosses and benefited from mechanization and mechanical detasseling in isolation nurseries, which are being replaced by natural or transgenic male sterility systems.

Inbred line development A major breakthrough in inbred line development has been the discovery of haploid plants and the concept of DH lines, which require only two instead of six or more generations to develop. Haploid plants are smaller and less vigorous than corresponding diploid plants (Chase 1952). Spontaneous parthenogenetic or androgenetic maize haploids occur at very low frequencies. Randolph (1940) discovered 23 parthenogenetic diploids among 17,165 individuals in the progeny of tetraploid maize, a frequency of about 1:750. Einset (1942) found two monoploids among 1916 plants, a frequency of 1:958. Chase found 43 monoploids among 38,684 seedlings using a dominant gene for purple plumule (Chase 1949). Stadler (1949) obtained a frequency of about 1:100 with a diploid multiple recessive tester. It took decades from the initial discovery of haploids to routine use of DH lines in maize breeding programs. Three main factors limited use of DH technology: low haploid induction rate, difficulty in the identification of haploid kernels and limited genome doubling capabilities. Besides haploid induction in vivo, in vitro techniques have been evaluated. However, maize turned out to be highly recalcitrant apart from a limited number of genotypes (Geiger 2009).

Haploid induction Chase (1949, 1951) suggested that haploids could be used for line development in hybrid breeding. Identification and use of haploid inducer Stock 6 was a major breakthrough for DH technology in the 1950s, with a maternal haploid induction rate of 2.3% (Stock 6 used as male). From the first use of inducer line Stock 6 to modern inducers, induction rates increased from 2% to close to 15% over about five decades. Rotarenco et al. (2010) reported the highest induction rate (14.5%) in their recently developed inducer lines derived from Stock 6 and MHI. At Iowa State University, induction rates above 15% were obtained for $F_{6:7}$ families from the cross of unrelated inducers with > 10% induction rates each (Frei, personal communication). We conclude that inducer lines with haploid-inducing capacity in excess of 20% are likely in the near future. Moreover, genes involved in maize haploid induction have been identified (e.g., Kelliher et al. 2017; Liu et al. 2017). A deletion in *Matrilineal* (*MTL*) causes significantly increased haploid induction rates in maize. Gene editing can thus be used to increase haploid induction rates in any genotype accessible to genome editing in maize.

Identification of haploid kernels About 90% of offspring from inducer crosses are regular diploids and thus undesirable. Coe and Sarkar (1964) developed several marker systems including R1-Navajo (*R1-nj*) and applied them to facilitate the identification of haploid seed. The *R1-nj* gene is

a dominant anthocyanin color marker gene, which expresses in the aleurone as well as the embryo. It enables the identification of kernels with haploid embryos based on (lack of) color. The *R1-nj* genetic marker that has to be sorted manually is still widely used. Expression of the *R1-nj* color marker can vary depending on genetic donor background and environmental factors (Liu et al. 2016). To overcome the shortcomings of *R1-nj*, alternative methods for automated sorting based on high oil, color, seed weight and near-infrared spectroscopic differences are under development (Liu et al. 2016). This includes the development of haploid inducers with high oil content, to facilitate haploid–diploid discrimination (Melchinger et al. 2015).

Genome doubling In the 1950s, colchicine was introduced to generate DH lines. Colchicine is an effective method for plant somatic genome doubling. Different protocols were established for maize Eder and Chalyk (2002). As colchicine is toxic to humans and the environment, alternative chemicals such as herbicides, such as nitrous oxide and trifluralin, have been proposed for genome doubling (Kato 2002; Häntzschel and Weber 2010).

Haploids may become fertile spontaneously by haploid genome doubling (SHGD). Barnabás et al. (1999) reported that SHGD rates ranged from 0 to 21.4% among maize germplasm. After *in vivo* haploid induction and planting in the field, maize haploids usually show a high degree of haploid female fertility (HFF) (Chase 1952; Chalyk 1994; Kleiber et al. 2012). More than 90% haploid ears with kernels are obtained after crossing haploid plants with regular diploid maize pollen (Chalyk 1994; Geiger et al. 2006). The average haploid male fertility (HMF) rate is usually below 10% (Chase 1949; Chase 1952; Chalyk 1994) which limits the number of DH lines produced in a population without colchicine treatment. Thus, methods to improve HMF are of interest to maize breeders. Geiger and Schönleben (2011) found significance within population variation for HMF, corroborated in studies of temperate and tropical germplasm (Zabirova et al. 1993; Chalyk 1994; Kleiber et al. 2012). Major QTL affecting HMF has been identified (Ren et al. 2017).

Future DH breeding schemes Further improvement in DH technology will reduce costs for inbred line development considerably. Breeding strategies that make best use of breeder genetic, technical and monetary resources have been proposed (Gordillo and Geiger 2008; Geiger 2009). Major breeding programs combine DH technology with genomic selection (GS) (see below), to maximize genetic gain per breeding cycle. With doubling rates exceeding 17%, the costs for GS at the haploid stage would be lower than conducting GS one generation later, at the diploid stage (Wu et al. 2015). Thus, if SHGD works well, maize breeding programs using both DH technology and GS can be accelerated.

Controlled pollinations With the first commercial hybrid seed production in 1923, manual detasseling of seed parents was employed to maximize hybrid purity and to avoid hand pollinations in hybrid seed production fields (Wych 1988). Manual detasseling has later been advanced to mechanical detasseling, or a combination of mechanical followed by manual detasseling for control (Wych 1988). Manual and mechanical detasseling remains an important method of hybrid seed production today; its use depends on the availability of alternative biological mechanisms, which allow hybrid seed production at lower costs. Manual detasseling was increasingly replaced by the use of cytoplasmic male sterility (cms) in the 1950s to 1970s, but gained renewed importance with the advent of Southern corn leaf blight, which eliminated the use of T-cytoplasm as a primary cms source for hybrid seed production.

cms in maize was first described by Rhoades (1931). Three major sources of cms have been recognized: cms-T (Texas), cms-C (Charrua) and cms-S (USDA) (Gabay-Laughnan and Laughnan 1994). While cms is caused by defects in mitochondrial DNA and, thus, maternally inherited, fertility in hybrids needs to be restored. This is accomplished by crossing cms females with males, carrying matching genic inherited restoration of fertility (Rf) genes. Rf1 and Rf2 restore the fertility of cms-T, Rf3 the fertility of cms-S, and Rf4 and Rf5 the fertility of cms-C (Gabay-Laughnan and Laughnan 1994). While actual seed production using cms is less costly compared to mechanical detasseling, both cms and Rf genes need to be introduced into the respective female and male parents, respectively. Moreover, cms and any biological systems for pollen control may be affected by environmental conditions. Fertility of cms females has been observed under some conditions (Jugenheimer 1985), which leads to self-pollination of females and reduced yield of respective hybrid seed lots. Alternative genic male sterility and chromosomal–genic systems have been developed (Duvick 1965), but the majority of seed produced using biological pollen control has been based on cms systems (Jugenheimer 1985).

Combination of DH technology and cms conversion Paternal haploid induction in maize is mediated by the gene *ig1* (indeterminate gametophyte), which increases the frequency of haploids in its progeny (Kermicle 1969). Homozygous *ig1* mutants show several embryological abnormalities including egg cells without a nucleus. After fusion with one of the two paternal sperm cells, such an egg cell may develop into a haploid embryo possessing the maternal cytoplasm and only paternal chromosomes. In selected genetic backgrounds, the HIR ranges from 1% to 2% (Kermicle 1994). Because of

low frequency of haploids, this system is not widely used to derive DH lines. However, the *ig1/ig1* genetic stock can be useful for the conversion of an inbred line to its cms form. For this purpose, *ig1/ig1* inducer lines with various cms-inducing cytoplasms have been created (Pollacsek 1992; Schneerman et al. 2000).

Transgenic methods aiding seed production Male sterility and fertility restoration were among the first transgenic traits available, including the Barnase/Barstar system (Mariani et al. 1990). A team at Corteva Agriscience developed the seed production technology (SPT) system in maize (Wu et al. 2016), which has been deregulated by USDA APHIS in 2011 and is thus available for commercial hybrid seed production in maize. The maize SPT maintainer line is a homozygous recessive male sterile transformed with a SPT construct containing (1) a complementary wild-type male fertility gene to restore fertility, (2) an α-amylase gene to disrupt pollination and (3) a seed color marker gene. The sporophytic wild-type allele enables the development of pollen grains, carrying the recessive allele. Only half carry the SPT transgenes. Pollen grains with the SPT transgenes exhibit starch depletion resulting from expression of α-amylase and are unable to germinate. Pollen grains that do not carry the SPT transgenes are non-transgenic and are able to fertilize homozygous mutant plants, resulting in non-transgenic male-sterile progeny for use as female parents. Because transgenic SPT maintainer seeds express a red fluorescent protein, they can be detected and efficiently separated from seeds that do not contain the SPT transgenes by mechanical color sorting. Alternative systems have been or are being developed. Monsanto's Roundup Hybridization System (RHS) utilizes a transgenic maize trait (MON87427) that exhibits tolerance to glyphosate in all plant tissues except male reproductive tissues (Feng et al. 2014). Thus, genotypes carrying this event can be used as females, and pollen sterility can be induced by glyphosate application at flowering. The multicontrol sterility (MCS) system (Zhang et al. 2018) is based on the male sterility 7 (ms7) mutation in maize and uses color and herbicide tolerance to discriminate between male-sterile and fertile seeds.

Future It is likely that transgenic mechanisms (including those generated by genome editing) will increasingly be used to produce maize hybrid seeds to overcome the need (and costs) of detasseling. Primary concerns are (1) environmental stability of male sterility systems, i.e., ability to produce high-purity hybrid seed independent of environmental variation, and (2) regulatory acceptance of using transgenic hybrid seed production systems, which will likely differ substantially among countries.

Hybrid performance—hypotheses and prediction

Hybrid performance and heterosis played an important role in the history of maize breeding. Consequently, long-term research questions relate to the biological underpinnings of heterosis and on developing methods to predict the hybrid performance of various combinations of inbred lines.

Hypotheses regarding mechanisms of heterosis Observations of hybrid vigor in maize stretch back to Darwin who was the first to systematically describe the phenomenon (Darwin 1889). Darwin corresponded with Asa Gray at Harvard throughout his experiments who was a mentor to James Beal (Singleton 1941). Shortly after leaving Harvard for a position at the Michigan Agricultural Research Station, Beal conducted the first experiment in which one variety of maize was detasseled and then pollinated by another (Singleton 1941). Across multiple years of trials, crossed plants were consistently found to out-yield open-pollinated individuals (Singleton 1941). Starting in 1905, George H. Shull began a series of experiments at Cold Spring Harbor Laboratory in which lines of maize were self-pollinated. Inbreeding resulted in a marked decrease in plant vigor, whereas a dramatic rise in vigor was observed, when selfed lines were crossed. These results were published in two seminal papers that laid the groundwork for hybrid maize breeding (Shull 1908, 1909) and were supported by substantial further work on the effects of inbreeding by Edward East (1908).

Almost immediately, those observing the phenomenon of heterosis in maize proposed causal biological mechanisms. Shull (1908) and East (1908) posited superior performance in hybrids was caused by heterozygosity itself, which acted as a physiological stimulus. This explanation of heterosis is known as the overdominance hypothesis. In contrast, the dominance hypothesis, first developed by Davenport (1908) and then clearly articulated by Bruce (1910), attributes heterosis to the masking of effects of deleterious alleles by dominant or partially dominant alleles, with each inbred line providing its own complement of dominant, favorable alleles. While the dominance hypothesis was generally supported during the early portion of the twentieth century, a surge of support for overdominance grew in the 1940s leading up to the first Heterosis Conference at Iowa State University in the summer of 1950 (Gowen 1952). During this conference, it was generally concluded that the dominance hypothesis explains the loss of vigor due to inbreeding and subsequent recovery upon crossing, but is insufficient to explain the marked increase in vigor of hybrids relative to open-pollinated varieties, which could only be explained by overdominance (Crow 1999). Following the Heterosis Conference, the pendulum shifted again toward support for the dominance hypothesis with accumulating evidence for both high mutation rates requiring complementation of

deleterious alleles in maize and for improved performance over time of inbred lines *per se*, presumably due to purging of deleterious alleles (Crow 1998). In addition, examples of a previously proposed mechanism, pseudo-overdominance (Jones 1917), were discovered, in which favorable, dominant alleles found in repulsion phase in a particular linked chromosomal region led to a signal that could be mistaken for overdominance. For instance, Moll and colleagues (1964) found that signatures of statistical overdominance in early cycles of selection quickly disappeared, with later cycles characterized by dominance, presumably due to decreasing linkage. Pseudo-overdominance was clearly illustrated when Stuber and co-authors found that a QTL for heterosis fractionated during fine mapping into two dominant QTL in repulsion phase (Stuber et al. 1992; Graham et al. 1997).

While dominance and overdominance have been the primary explanations for heterosis, epistasis has often been described as an additional mechanism. However, epistasis appears to play a minor role in heterosis in maize (Melchinger et al. 1986; Garcia et al. 2008), but may be more important in self-pollinating, homozygous species such as rice in which dominance is likely less pervasive and epistatic interactions may be more stable (Garcia et al. 2008).

More recently, our understanding of heterosis in maize has been informed by genomic data. For example, with complex, quantitative traits, genomic data have linked heterosis to the combined effects of a number of loci (Stuber et al. 1992; Giraud et al. 2017). Separate traits (e.g., yield, plant height) have demonstrably unique genetic architectures of heterosis in the same hybrid-parent triplet, confirming a multigenic nature of heterosis (Flint-Garcia et al. 2009). Genomic data have also led to an elaboration of existing hypotheses regarding the mechanisms of heterosis. Comparison of sequences in different maize inbreds revealed a surprising level of presence/absence variation (PAV) in which sequences found in one inbred are lacking in another. For example, investigation of the *bz* locus in maize (Fu and Dooner 2002; Wang and Dooner 2006) revealed that inbred lines share only 50% of their sequence in this chromosomal region. These findings suggest that dominance, previously attributed solely to complementation of slightly deleterious alleles, may also involve complementation of absent sequence. Recently, Baldauf et al. (2018) had demonstrated that many more genes were actively expressed in hybrids than in their inbred parents. In several instances, this was due to absence of the gene in a particular inbred; in many more cases, the gene was present in the inbred but not expressed. Such scenarios could lead to expression complementation. Similarly, recent work has shown that dysregulation of expression (i.e., aberrantly low or high levels of expression of a given gene) can be caused by deleterious alleles (Kremling et al. 2018). At such loci, hybrids may experience mid-parent values of expression within an optimal range resulting in an increase in fitness in hybrids relative to the inbred parents as proposed by Springer and Stupar (2007).

While accumulating evidence in the molecular and genomic era continues to favor dominance as the prevailing mechanism driving heterosis in maize, the phenomenon is observed in highly quantitative traits that interact in complex pathways to produce a given phenotype. In all likelihood, diverse mechanisms including overdominance and epistasis play at least a minor role in heterosis and a single, unifying mechanism cannot be determined (Kaeppler 2012). A further complicating factor has been the observed trend of decreasing percentage of heterosis over time (Fig. 5), with a concomitant increase in the yield performance of hybrids and parental inbred lines (Troyer and Wellin 2009). This finding may be linked to the continued purging of deleterious alleles among inbred lines with each heterotic group but the general separation between heterotic groups.

Development of methods to predict hybrid performance
Despite gaps in our understanding of the mechanism of heterosis, substantial progress has been made in predicting hybrid performance. As hybrid breeding programs became established, the number of inbreds within each heterotic group increased dramatically. Soon, it became unfeasible to phenotypically evaluate performance of all single-cross hybrids due to the overwhelming number of pairwise combinations of inbred lines. Unfortunately, the evaluation of inbred lines *per se* has proved to be an ineffective predictor of hybrid performance due to the prevalence of strong dominance effects (e Gama and Hallauer 1977; Smith 1986). Therefore, the focus in evaluation of hybrid performance has since shifted to model-based prediction using both pedigree and genomic data.

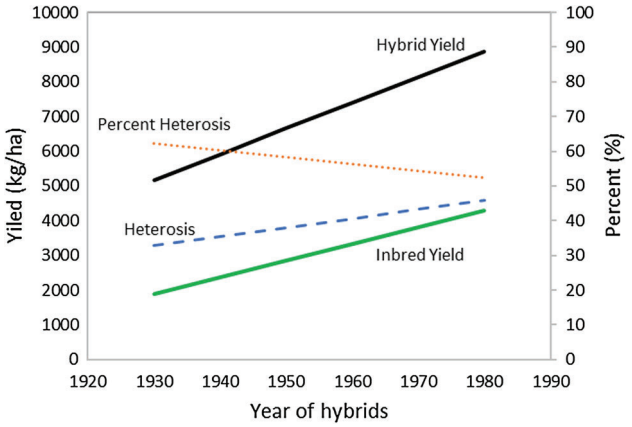

Fig. 5 Changes in hybrid yield, inbred yield, heterosis and percent heterosis along year of hybrids. Percent heterosis is calculated as heterosis over hybrid yield. Adapted from Troyer and Wellin (2009), Crop Science 49:1969–1976

In his pioneering work, Rex Bernardo modified the classical best linear unbiased prediction (BLUP) approach of Henderson (1975), predicting the performance of single-cross hybrids based on both yield data from related single crosses and a relationship matrix derived from molecular marker data from the parental inbreds (Bernardo 1994). This approach is now commonly known as genomic BLUP or GBLUP. In a subsequent study, Bernardo implemented this procedure to predict maize single-cross traits including yield, moisture, stalk lodging and root lodging in a population large enough to approximate a modern commercial breeding program (Bernardo 1996a). While prediction accuracies were high when both parents were tested in single-cross combinations, they dropped considerably when parents were not tested (Bernardo 1996b). Several statistical modifications and improvements upon the basic GBLUP approach have since been made (cf., Zhao et al. 2015; Desta and Ortiz 2014). For example, the ridge-regression BLUP approach (RR-BLUP; Whittaker et al. 2000) can predict the effects of individual markers on hybrid performance and Bayesian methods allow for a range of variances of individual marker effects (Zhao et al. 2015). Such methods have been implemented in maize using a process known as genomic selection (GS), in which high-density marker data are employed without pre-screening in order to determine genotypic values (Piepho 2009). GS has been shown to predict single-cross hybrid performance in maize at high accuracy even in germplasm from the early stages of a hybrid maize breeding program (Kadam et al. 2016). Quite recently, increasing attention has been paid to incorporating data into hybrid performance prediction that reflect intermediary steps between genotype and phenotype such as expression and metabolomic data (Schrag et al. 2018). Finally, genomic selection models have been shown to improve when modified to include annotation of deleterious alleles (Yang et al. 2017a, b).

Breeding project designs

A brief history The first modern maize breeders (George Shull, Edward East, Donald Jones, Henry Wallace, Perry Holden, Raymond Baker and George Sprague) were cognizant of evidence for response to selection provided by William Beal, Charles Darwin, Isaac Hershey, George Krug, Jake Leaming and Robert Reid (Kingsbury 2009) as well as the theoretical basis for response to selection (Fisher 1930). They designed breeding systems based on their research objectives, their understanding of heterosis and constraints imposed by reproductive biology and available resources. Due to its flexible reproductive biology, designs of maize breeding projects were more numerous than those developed for other crops (Comstock et al. 1949; Hull 1945; Jenkins 1940). Let us consider these designs from the perspective of two distinct objectives: to improve average population performance for a particular trait and to develop hybrids for sales to corn farmers.

Academic maize breeders designed recurrent population improvement projects using numerous locally adapted, open-pollinated (Leaming, Midland, Hays Golden, Golden Glow, Krug, Reid, Dawes, Iowa Ideal, Indian Chief, Jarvis, Burr's White, Lancaster, Kolkmeier) and synthetic populations (BS, BSSS, BSCB, CGSyn, EZS, NDS, VCBS). The purpose of these projects was to evaluate responses to selection in local environments using various types of selection units including: mass selection of individual plants, half-sib family selection, full-sib family selection and self-pollinated family selection. These same selection units also were evaluated for performance in hybrid combinations using reciprocal recurrent selection projects. Depending on the genetic variability in the founder populations, a wide range of responses were observed for all of the methods (Hallauer et al. 2010). Some of the recurrent selection projects were coupled with line development projects (Fig. 6a) that occasionally produced exceptional lines used in production of hybrids broadly grown by farmers: e.g., B13, B37, B73, B84.

The basic design of hybrid maize development pipelines (Fig. 6b) was well established by the 1970s. The goals of these projects were to: (1) maximize additive genetic variance and minimize contributions from non-genetic variance through the development of replicable homozygous lines within heterotic groups (Eberhart 1970); (2) evaluate lines *per se* for parental attributes and in hybrid combinations for agronomic attributes from two or more heterotic groups using replicated small plot field trials; (3) improve breeding lines genetically within each heterotic group by recycling lines with desirable agronomic attributes; (4) identify the best hybrid-environment combinations for selected hybrids using large-scale, on-farm, field trials requiring practical aspects of preparing foundation, registered and certified seed (Fehr 1991).

The predicted response to selection or predicted genetic gain for each cycle of breeding has several forms:$\Delta G_c = \iota\sigma_p h^2 = \iota\sigma_a h = \iota\sigma_a \rho$. For maize breeders, the predicted genetic gain per year, $\Delta G_t = \Delta G_c/\text{years}$ (Eberhart 1970; Hallauer and Miranda 1981), has been a metric for making decisions about proposed modifications to breeding methods (Fehr 1991). Based solely on the criterion of cycle time, population improvement projects in the 1970s were faster than hybrid development projects. They intermated selected lines every three to 5 years, while line and hybrid development projects intermated selected lines every seven to 10 years. The advantages of the pipelines model are that they provide opportunities for evaluation and selection across years (stages) which enabled greater selection intensities, ι, for multiple traits and they produced greater correlations, ρ, between selection units and response units (Holland

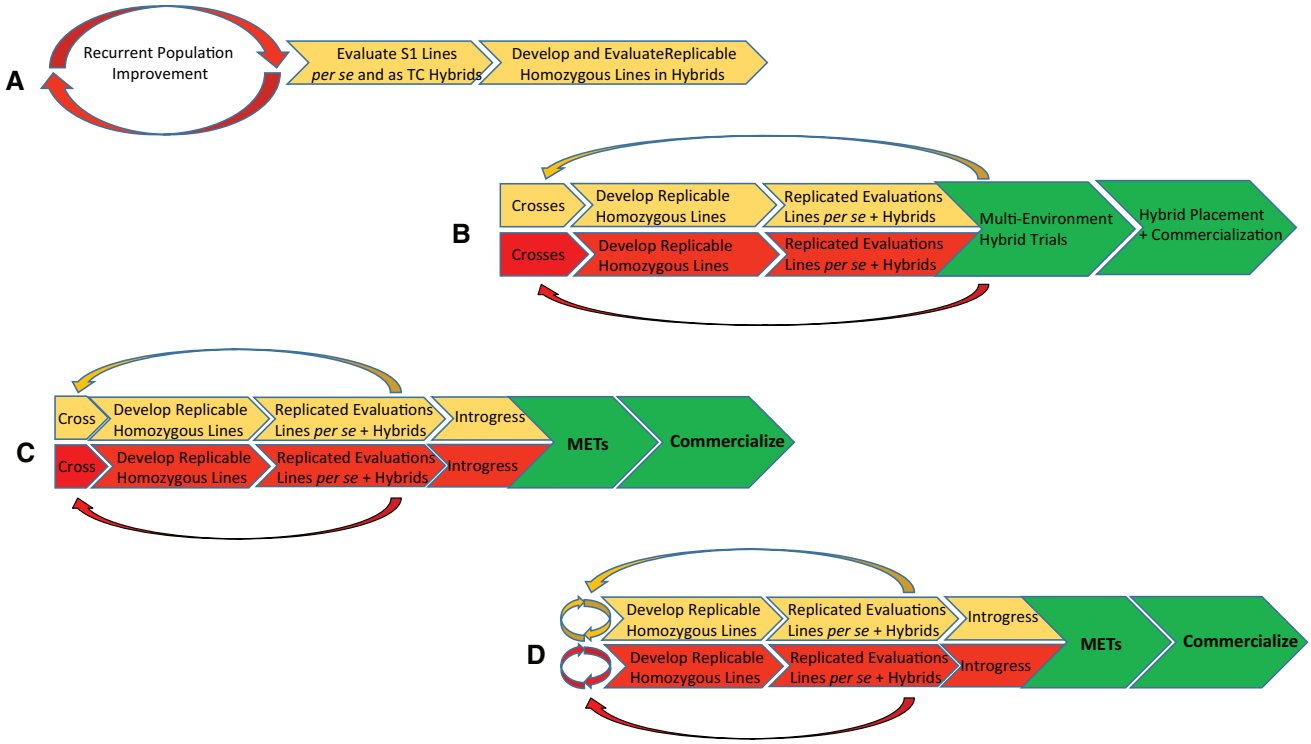

Fig. 6 **a** Depiction of relationship between recurrent population improvement projects and line development projects; **b** depiction of maize hybrid development as consisting of parallel line development pipelines (red and yellow) within heterotic groups and a hybrid evaluation and commercialization pipeline (green). Lines advanced to late stages with desirable attributes are used in crossing nurseries to recurrently initiate the development of novel replicable lines; **c** depiction of maize hybrid development pipelines modified to include trait introgression within heterotic groups; **d** depiction of maize hybrid development pipelines modified to include introgression of non-negotiable traits for hybrid sales and rapid cycling through genomic selection for population improvement. Adapted from (Gaynor et al. 2017) (color figure online)

et al. 2003). Perhaps more importantly, the pipeline models were designed to develop, evaluate and disseminate hybrids to farmers, i.e., they were economically sustainable. Indeed, many land-grant universities in the USA initially supported line and hybrid development projects, but only a few remain because public maize breeding did not capture the value of their released lines and hybrids in the marketplace. Currently most maize breeding projects in the USA and Europe are supported by revenues from sales of commercial seed. There is significant effort on behalf of nonprofit organizations to implement maize line and hybrid pipelines in an economically sustainable manner for developing countries (Gary Atlin, personal communication).

The basic form of the maize hybrid development pipelines has been sufficiently robust to incorporate a large number of technological innovations such as expanded evaluation phases for germination tests, and disease and insect nurseries to protect genetic gains (Dicke and Guthrie 1988; Smith and White 1988). In the 1990s, the pipelines became longer with the introduction of new pipeline segments to accommodate marker-assisted introgression of transgenic events from poorly adapted, but transformable, lines (Fig. 6c).

Backcross-enabled introgression has been practiced for a long time (Johnson and Eldredge 1953; Wilcox and Cavins 1995), and marker-assisted introgression will likely continue for breeding teams that have not negotiated access to enabling technologies for maize genome editing (Lowe et al. 2016) or lack sufficient resources to address regulatory requirements. Potential target alleles for introgression have been discovered using both forward and reverse genetic approaches in germplasm resources and gene banks (Blumel 2015; Kumar et al. 2010; Leung et al. 2015) and have been catalogued in MaizeGDB (Lawrence et al. 2008). Emerging high-throughput technologies have been proposed to increase the pace of genetic discoveries (Yu et al. 2016), although the "turbocharged" discovery process will require the development of automated curation processes for MaizeGDB to keep pace.

While some technological innovations added components to the pipelines, other technological innovations were adopted to reduce time required to develop new hybrids and coincidently reduce time per breeding cycle (Li et al. 2018). For example, the development of "winter nurseries" in tropical locations and subsequent development of continuous

nurseries in tropical and high-latitude locations reduced the time to develop replicable homozygous lines. The time to develop replicable homozygous lines was further reduced with the development of doubled haploid capabilities. Also the number of years required for hybrid field evaluations was reduced through the use of locations at equivalent latitudes in opposite hemispheres (Cooper et al. 2014). Continuous nurseries and continuous field trial evaluations required information technologies, logistical software systems and development of high-throughput seed handling and transfer capabilities to assure timely tracking and delivery of seed (Serhatli et al. 2018). As a consequence, the time required to develop and deliver a new hybrid was reduced from more than a dozen years in the 1970s to about 7 years, while the time required per breeding cycle was reduced from about 10 years to as little as 5 years by the end of the first decade of the twenty-first century. The associated costs for implementing these innovations are not clear, not even for stockholders.

Discoveries of genome organization, maize diversity, genetic signaling networks and metabolic pathways, as well as development of digitized phenomics, precision envirotyping (Xu 2016), genomic selection (Cooper et al. 2014; Meuwissen et al. 2001), genome editing (Shi et al. 2017; Svitashev et al. 2015) and speed breeding (Watson et al. 2017) coupled with eco-physiological crop models (Technow et al. 2015) have potential to modify or completely redesign maize genetic improvement models. A comprehensive descriptive review of how these discoveries and technical innovations could affect the components of ΔG_t was provided by (Xu et al. 2017). As with pre-molecular breeding, the overriding theme of most is to reduce time per cycle of breeding. With the exception of genomic selection, reductions in time have been associated with increased costs. A new conceptual framework is needed to address the trade-offs between time and cost because corn growers are questioning the value of continuously escalating costs for seed and other inputs (Serhatli et al. 2018).

Exploring possible modifications to maize breeding projects The current ratio of annual genetic gains to seed costs is not sustainable, as evidenced by global consolidation of breeding and ag chemical organizations. As a consequence, maize breeding programs will be expected to modify or redesign their development pipelines to double the rates of genetic gain with fewer resources (lower costs). For the last 20 years, suggested modifications to maize (plant) breeding pipelines have been investigated using simulations rather than descriptive assessments of impacts on the components of ΔG_t. Simulations are used because the functional relationships for annual genetic gains of a single-trait, as well as its multiple-trait version, $\Delta H_c = ir_{IH}\sigma_H = \sum a_i \Delta G_{c_i}$ (Hazel 1943; Smith 1936), depends on simplifying assumptions about the genetic architecture (infinitesimal additive model), genome organization (no linkage), population structure (Hardy–Weinberg equilibrium) and nature of selection intensity (single-stage truncation).

Since none of the assumptions underlying the breeder's equation are correct for any specific plant breeding project, computer simulations using models for genetic architectures, genetic linkage, population structures, multistage selection and assortative mating were developed (Cooper and Podlich 2002; Fraser and Burnell 1970; Peccoud et al. 2004; Tinker and Mather 1993) to explore specific impacts of modifications to breeding pipelines (Cress 1967; Li et al. 2012a; Mi et al. 2014; Podlich and Cooper 1998; St Martin and Skavaril 1984; Sun et al. 2011). The initial simulation studies focused on discoveries about nonadditive genetic architectures and dependencies on environmental signals (Cooper et al. 2002; Cooper and Podlich 2002; Peccoud et al. 2004) or resource allocations in terms of numbers of plots and environments (Longin et al. 2007; Wang et al. 2004, 2007). Also, simulation experiments have been reported for marker-assisted backcross introgression (Cameron et al. 2017; Chevalet and Mulsant 1992; Frisch et al. 1999; Herzog and Frisch 2011; Herzog et al. 2014; Hillel et al. 1990; Hospital 2001; Hospital and Charcosset 1997; Peng et al. 2014a, b; Visscher et al. 1996), accuracy of genomic prediction methods (Daetwyler et al. 2010; Heslot et al. 2012; Howard et al. 2014), genomic selection across stages within line and hybrid development pipelines (Longin et al. 2007, 2015; Marulanda et al. 2016; Mi et al. 2014) and genomic selection across stages and cycles (Bernardo and Yu 2007; Gaynor et al. 2017; Gorjanc et al. 2018; Heffner et al. 2009; Jannink 2010).

Most simulation studies can be characterized as exploratory *in silico* experiments. They not only allow the investigators to accommodate deviations to the assumptions underlying the theoretical breeder's equation, but also allow the investigators to explore replicated combinations of factors and modifications to pipelines that could affect outcomes (Gaynor et al. 2017; Podlich and Cooper 1998; Sun et al. 2011). A clear advantage of this approach is that sample sizes, numbers of replications, numbers of factors and modifications can be very large, and the time required to conduct an *in silico* experiment is very small, thus allowing the investigators to screen a large number of possible designs before investing in possible expensive and time-consuming product development projects.

As with all experimental approaches, simulation experiments are subject to an investigator bias that constrains the consideration of possible factors and modifications to development projects. Nonetheless, an innovative design has emerged for integrated recurrent genomic selection and conventional line development projects (Gaynor et al. 2017). The innovation (Fig. 6d) is referred to as a two-part program consisting of a conventional product development component that develops and screens inbred lines using established

pipelines and a genetic improvement component enabled by rapid cycling with recurrent genomic selection. Models used in the population improvement part are created using training sets consisting of phenotypic and genotypic data from established lines. Furthermore, drift between the rapid cycling genetic improvement part and the training set from development pipelines as well as the loss of genetic diversity through selection can be reduced by selecting crosses that balance the trade-offs between maintaining genetic diversity and genetic gain (Gorjanc et al. 2018).

Future maize breeding teams, however, should be careful to avoid confusing outcomes from exploratory *in silico* experiments with designing projects for specific breeding objectives. Exploratory *in silico* investigations include discovery objectives. In contrast, a systematically designed project will explicitly state objectives in terms that can be optimized.

An engineering approach to design optimal maize breeding projects Merriam-Webster (https://www.merriam-webster.com/) considers optimization to be the process of designing a system to be the most effective and efficient possible, as determined by mathematical models and computational procedures. Operations research (OR) is a subdiscipline of applied mathematics devoted to the study of optimization of complex systems. OR was created to provide quantitative risk assessments of proposed military activities for uncertain and dynamic conditions in WWII and has since been used to design optimal manufacturing, transportation, energy and communications systems and networks. Some of the first civilian applications were in agriculture (Boles 1955; Heady 1954; Heady and Pesek 1954; Rendel and Robertson 1950; Robertson 1957), but with one exception (Johnson et al. 1988) OR was ignored for designing plant breeding systems until about 10 years ago (Akdemir et al. 2018; Akdemir and Sanchez 2016; Byrum et al. 2016, 2017; Cameron et al. 2017; Canzar and El-Kebir 2011; De Beukelaer et al. 2015; Han et al. 2017; Xu et al. 2011).

OR approaches are based on systematic development of mathematical programming (MP) models. MP models are comprised of: clear measurable *objectives* that are translated into *objective functions* consisting of *decision variables* and *constraints* (Winston et al. 2003). Often objectives are not complementary, and the optimization problem consists of multiple competing objectives such as minimizing time while assuring the probability of success is > 0.95 for an enforced budget constraint. In a MP model, the components of annual genetic gain, such as described in Table 1 of (Xu et al. 2017), would be considered decision variables, while the "subcomponents" and "contributors" would be considered constraints. Decision variables are parameters of the objective functions that can be controlled, e.g., numbers of years to complete a cycle, number of progenies to grow per evaluation stage, number of markers or phenotypes to assay and possible selection intensities. Constraints are limitations on the decision variables, e.g., budget restrictions, time required for stages of phenology, reproductive biology, assay deadlines, size and number of field plots.

After translating the objectives into mathematical functions, and the decision variables and constraints into a MP model, the model may be recognized as belonging to a family of models that have been previously solved analytically. For example, an optimization model to maximize annual genetic gains is likely to be of the same form as nonlinear programming models (Winston et al. 2003). If it is, then there exists a large library of algorithms to solve nonlinear programming models some of which use KKT conditions (Karush 1939; Kuhn and Tucker 1951) providing a set of feasible solutions that have been proved to be among the best possible. Most of these algorithms have been implemented in solver software packages such as GUSEK, R, MATLAB and even EXCEL.

The allure of analytic solutions supported by mathematical proofs to find a best possible breeding design should be tempered by recognizing that optimization MP models for plant breeding projects will seldom consist of a single objective, nor will there be known functional relationships among constraints because stochastic processes rather than mechanistic functions generate envirotypes, annual budgets and genotype to phenotype relationships. The relationships among these and their impact on meeting the objectives will need to be investigated with simulations similar to those used for exploratory *in silico* experiments.

Examples of optimal designs for plant breeding projects As previously noted, there are many published exploratory simulation experiments for backcross introgression of a single allele from homozygous donors to homozygous recipients (Frisch et al. 1999; Herzog and Frisch 2011; Herzog et al. 2014; Hospital 2001; Hospital and Charcosset 1997; Hospital et al. 1992; Peng et al. 2014a, b). From among these publications (Peng et al. 2014a, b) investigated the largest number of backcross breeding strategies, where each strategy represented a combination of backcrossing generations, number of progeny per generation and numbers and genomic distributions of markers assayed per backcross generation. They identified one strategy that was better than all others with respect to recovering the introgressed allele, the average recovery of the recipient genome among four selected progenies, and minimal number of backcross generations, minimal number of progeny and minimal marker costs. Rather than exploring a larger number of possible backcrossing strategies, Cameron et al. (2017) formulated an optimization model in which the objective was to maximize the probability of success, where success was clearly defined for a genome with the desirable allele in a homozygous condition

and no more than the average donor genome described for the best strategy identified by Peng et al. (2014a, b). The solution to the optimization model was a backcross design that doubled the probability of success for about the same cost and time as that of the best model found by Peng et al. (2014a, b). Subsequently, the optimization model was modified by removing the constraint of backcrossing selected progeny to the recipient line every generation. The best solution to the modified model increased the probability of success by another 20% and reduced costs and time. In other words, backcrossing is not always the best action to pursue in an introgression project (to be published at a later date). This also illustrates that changing constraints and/or decision variables of the optimization model will likely produce a different (often better) design to meet the objectives.

Introgression of a single desirable allele inevitably produces the challenge to design projects that combine several desirable alleles from multiple donors into a single recipient line, a.k.a. gene stacking. Initially possible gene stacking designs were suggested based on knowledge of inheritance in the context of traditional breeding designs (Ishii and Yonezawa 2007; Peng et al. 2014a, b; Servin et al. 2004; Ye and Smith 2008). Because gene stacking is similar to assembling products in a manufacturing system, it has been amenable to mathematical programming approaches (Canzar and El-Kebir 2011; De Beukelaer et al. 2015; Xu et al. 2011). Since every gene stacking problem is unique, there will be no single best solution; rather, the solution will need to be found for each set of alleles and the distributions of their specific genomic locations as well as the distributions of the marker loci. Nonetheless, an OR approach should be capable of developing a MP model that can be solved.

After a set of desirable alleles has been stacked into a single line, there will be a need for an optimal design to transfer these to other lines in a product development project. This type of objective might also arise when attempting to transfer multiple desirable alleles from one germplasm group to another. For example, consider the evaluation of tropical lines in high latitudes as a new source for useful genetic variability. Recall that most tropical lines are photoperiod-sensitive. If photoperiod-insensitive alleles from lines adapted to high latitudes were transferred to the tropical lines, the genetic and geographic barriers for genetic improvement would be lowered as it has been in Sorghum (Klein et al. 2008, 2016). There are between a dozen and two dozen loci (Buckler et al. 2009; Romero Navarro et al. 2017) that will need to be introgressed from high-latitude maize lines to produce photoperiod-insensitive tropical introgression lines. Genomic selection is a reasonable approach for this type of challenge (Bernardo 2009). By framing multiallelic introgression as an optimization model, (Han et al. 2017) realized the need for a metric other than genomic estimated breeding values (GEBVs) that would assign a specific combining ability, known as the predicted cross value (PCV), to all possible crosses among the sample of progeny created each generation. In contrast, the estimated GEBV is analogous to general combining ability. A comparison of the PCV with GEBVs and optimized haploid values (OHVs) revealed that selecting specific crosses based on the PCV, rather than random crosses among truncation selected individuals with high GEBVs and OHVs, produced the desired introgression lines in less time (Han et al. 2017). Optimal designs for the use of the PCV are still being developed for introgression projects (submitted). Also MP models have been used to increase the efficiencies of genomic selection for individual and multiple traits by enabling selection of individual crosses instead of truncation selection based on GEBVs (Akdemir et al. 2018; Akdemir and Sanchez 2016).

The power of genomic selection to increase genetic gains through greater selection intensities and shorter time intervals, like all forms of intense selection, could exhaust useful genetic variability and prevent plant breeding populations from reaching their full genetic potential (Bulmer 1971; Hill and Robertson 1968; Jannink 2010; Robertson 1960). Animal breeders have been using MP to address the challenge of minimizing the trade-offs between genetic gain and loss of genetic diversity in breeding populations for at least 20 years (Eynard et al. 2018; Fernandez and Toro 1999; Howard et al. 2017; Kinghorn 1998; Pong-Wong and Woolliams 2007; Woolliams et al. 2015). To our knowledge, MP models still need to be developed to address this challenge in designing maize (plant) genetic improvement and hybrid development pipelines.

Perhaps the most impactful application of OR to plant breeding began in 2009 when Syngenta teamed with Kromite, an OR and decision analytics company, to evaluate the efficiency of breeding systems inherited from the merger of three companies (Byrum et al. 2016). They first recognized that genetic improvement, trait introgression, variety development and variety placement were distinct projects within their breeding programs. Next, they disentangled the variety development pipelines from genetic improvement, trait introgression and product placement projects. They also had distinct variety development pipelines for each maturity group. Overall they identified about 250 decision points per year for variety development. If the decisions were independent and binary (they are not), there would be at least 2^{250} possible outcomes from the existing pipelines. Next, the evaluation of possible outcomes required critical thinking about appropriate metrics to quantify impacts on meeting their breeding objectives. Based on the development of novel metrics (Byrum et al. 2017) and a comprehensive exploration of the modification space for each variety development project, they implemented modified variety development pipelines, resulting in over $287 M US cost savings during the period 2010 to 2015 and awarding of the 2015 Edelman

prize (https://www.informs.org/About-INFORMS/News-Room/Press-Releases/Syngenta-Wins-2015-INFORMS-Edelman-Prize). To our knowledge this approach has not been applied to maize hybrid development pipelines, but given the pressure to reduce costs, it is likely that several commercial maize breeding companies are pursuing similar approaches to quickly become more efficient.

Challenges of the OR approach for maize (plant) breeders The most difficult aspect of executing the OR approach is to clearly define the objectives so that they can be translated into MP models. This is nothing new because the most difficult aspect of maize breeding has always been to clearly define objectives in terms of measureable metrics.

With the development of large databases containing historical field trials, historical weather records and the ability to merge data from these databases, it has been possible to develop envirotypes that are associated with stable and plastic responses (Li et al. 2018; Xu 2016). Linking envirotyping with crop growth models, Cooper et al. (2014) demonstrated that measurable breeding objectives for targeted environments can be clearly articulated, and more importantly, by enlisting precision digital phenotyping and genomic selection the objectives were met and a drought-tolerant hybrid was delivered to the marketplace. Interestingly, the same objectives were met using genome editing by a different research group at the same company (Shi et al. 2017). While both genomic selection and genome editing effectively developed hybrids that were drought-tolerant, it is not clear which approach is more efficient. Also, given the relatively small number of genome edits required to create drought resistant hybrids, there may be genetic introgression approaches that are more efficient.

The question of which approach is most efficient will depend, in part, on whether the breeding team has access to enabling technologies for genome editing and sufficient resources to address regulatory requirements, but more importantly what metrics will represent a successful outcome? How much time and capital will optimally achieve the breeding objectives? These questions frame the optimization challenge in terms of cost, time and probability of success (CTP). With competing objectives in the CTP framework, there will not be a single best solution. Instead, there will be a set of optimal solutions, a.k.a., Pareto frontiers that help decision makers to quantify trade-offs in which it is not possible to improve one objective without degrading others.

While it is tempting to frame optimization models using a CTP recipe, defining success needs to be based on predicted benefits. Success framed in terms of forecast benefits will enable quantification of the trade-off between time and cost. For example, more expensive, shorter times to complete a project could bring greater benefits by being first to market to offset increased costs. Unfortunately, plant breeders and agronomists, in general, have little experience with developing models to forecast benefits, especially in terms of net present value. Forecasting benefits, like predicting phenotypes, is based on uncertain outcomes from stochastic processes. Fortunately, OR has been successfully integrating forecasts and risk assessments for optimal outcomes involving uncertainty since its creation in WWII (Birge and Louveaux 2011). We suggest that it is time for maize (plant) breeding teams to return to their original role as designers of systems for genetic improvement, line and hybrid development, introgression and product placement by learning OR approaches and collaborating with experts in stochastic programming.

Synopsis

Maize and generally plant breeding is gradually transitioning from a black box approach, largely agnostic of genes and alleles affecting trait variation, to a discipline, where decisions are based on a combination of deep understanding of which combinations of genes and respective alleles lead to improved breeding populations and cultivars and massive testing of selected candidates based on prior performance and predictions. Currently, genomic selection is an extension of traditional selection methods, where large numbers of genotypes are evaluated first at DNA and a selected fraction at the more expensive agronomic levels, to ultimately identify a limited number of superior experimental variety candidates. With a more complete understanding of which gene and respective allele combinations would result in the optimal genotype for a given environment, more targeted approaches are expected to emerge, which will enable their design. Hypothetically, such designer genotypes could be obtained using traditional recombinant technology (by crossing carefully selected founder genotypes), or by editing multiple genes. However, there remain challenges in obtaining a more complete understanding of trait variation, including the classical challenges of small genetic effects of QTL, difficulty in predictions in the presence of complex G x G and G x E interactions. Nonetheless, the breeder toolbox is becoming populated with modernized tools. The question is no longer, whether a tool or approach is available, but which of the increasing number of options should be chosen, given limited financial resources, to maximize both genetic gain and economic return. Availability of genome editing as a powerful tool for plant breeders will be substantially affected by its regulatory framework. Even if liberal in some countries, its use may be limited by more restrictive regulations in other countries. Changing climates may drive the need to use broader genetic resources, which in the longer run may help not only to close the gap between actual and potential yields, but also to raise the bar for potential yields.

Author contribution statement All authors (CA, WDB, MH, SS, WS, KW, MW, JY, TL) contributed equally to design, writing, and revision of this review article. TL served as corresponding author to assemble drafts provided by individual co-authors, to submit files to TAG, and to finalize the submission process.

Acknowledgements The authors would like to thank USDA's National Institute of Food and Agriculture (Project numbers: IOW04314, IOW05520), as well as the RF Baker Center for Plant Breeding at Iowa State University for supporting this work.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Akdemir D, Sanchez JI (2016) Efficient breeding by genomic mating. Front Genet 7:1–12

Akdemir D, Beavis W, Fritsche-Neto R, Singh AK, Isidro-Sánchez J (2018) Multi-objective optimized genomic breeding strategies for sustainable food improvement. Heredity. https://doi.org/10.1038/s41437-018-0147-1

Amano E, Smith HH (1965) Mutations induced by ethyl methanesulfonate in maize. Mutat Res 2:344–354

Anderson E, Cutler HC (1942) Races of *Zea mays*. I. Their recognition and classification. Ann Mo Bot Gard 29:69–89

Andorf CM, Cannon EK, Portwood JL 2nd, Gardiner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalusu VV, Huerta M, Cho KT, Wimalanathan K, Richter JD, Mauch ED, Rao BS, Birkett SM, Sen TZ, Lawrence-Dill CJ (2016) MaizeGDB update: new tools, data and interface for the maize model organism database. Nucleic Acids Res 44:D1195–D1201

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

Barnabás B, Obert B, Kovács G (1999) Colchicine, an efficient genome-doubling agent for maize (*Zea mays* L.) microspores cultured in anthero. Plant Cell Rep 18:858–862

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-Gonzalez J, Ouzunova M, Revilla P, Charcosset A, Martin OC, Schön CC (2013) Intraspecific variation of recombination rate in maize. Genome Biol 14(9):R103

Beavis WD, Grant D (1991) A linkage map based on information from four F_2 populations of maize. Theor Appl Genet 82:636–644

Beckett TJ, Morales AJ, Koehler KL, Rocheford TR (2017) Genetic relatedness of previously Plant-variety-protected commercial maize inbreds. PLoS ONE 12(12):e0189277

Bedoya CA, Dreisigacker S, Hearne S, Franco J, Mir C, Prasanna BM et al (2017) Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. PLoS ONE 12(4):e0173488

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58:268–276

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. Nucleic Acids Res 41:D36–D42

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25

Bernardo R (1996a) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Bernardo R (1996b) Best linear unbiased prediction of the performance of crosses between untested maize inbreds. Crop Sci 36:872–876

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49:419–425

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Betran FJ, Ribaut JM, Beck D, Gonzalez de Leon D (2003) Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and nonstress environments. Crop Sci 43:797–806

Birchler JA (1980) The cytogenetic localization of the alcohol dehydrogenase-1 locus in maize. Genetics 94:687–700

Bird RM, Neuffer MG (1987) Induced mutations in maize. In: Janick J (ed) Plant breeding reviews. Van Nostrand Reinhold, New York, pp 139–180

Birge JR, Louveaux V (2011) Introduction to stochastic programming. Springer, New York

Boles JN (1955) Linear programming and farm management analysis. J Farm Econ 37:1–37

Bolser DM, Staines DM, Perry E, Kersey PJ (2017) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. Methods Mol Biol 1533:1–31

Bommert P, Nagasawa NS, Jackson D (2013) Quantitative variation in maize kernel row number is controlled by the *FASCIATED EAR2* locus. Nat Genet 45:334–337

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):e71377

Brandenburg J-T, Mary-Huard T, Rigaill G, Hearne SJ, Corti H, Joets J, Vitte C, Charcosset A, Nicolas S, Tenaillon M (2017) Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLoS Genet 13(3):e1006666

Brown WL, Goodman MM (1977) Races of corn. In: Sprague GF (ed) Corn and corn improvement. Amer Soc Agron, Madison, pp 49–88

Brown AHD, Hodgkin T (2015) Indicators of genetic diversity, genetic erosion, and genetic vulnerability for plant genetic resources. In: Ahuja MR Jain SM (eds) Genetic diversity and erosion in plants, sustainable development and biodiversity vol 7, pp 25–53

Bruce AB (1910) The Mendelian theory of heredity and the augmentation of vigor. Science 32:627–628

Brunelle DC, Clark JK, Sheridan WF (2017) Genetics screening for EMS-induced maize embryo-specific mutants altered in embryo morphogenesis. G3 7:3559–3570

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, Sanchez Villeda H, da Silva HS, 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 325:714–718

Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, Wang B, Xu D, Yang B, Xie C, Fan L, Gao S, Xu X, Zhang G, Li Y, Jiao Y, Doebley JF, Ross-Ibarra J, Lorant A, Buffalo V, Romay MC, Buckler ES, Ware D, Lai J, Sun Q, Xu Y (2018) Construction of the third-generation *Zea mays* haplotype map. GigaScience 7:1–12

Bulmer MG (1971) The effect of selection on genetic variability. Am Nat 105:201–211

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Byrum J, Davis C, Doonan G, Doubler T, Foster D, Luzzi B, Mowers R, Zinselmeier C, Klober J, Culhane D, Mack S (2016) Advanced analytics for agricultural product development. Interfaces 46:5–17

Byrum J, Davis C, Doonan G, Doubler T, Foster D et al (2017) Genetic gain performance metric accelerates agricultural productivity. Interfaces 47:442–453

Cameron JN, Han Y, Wang L, Beavis WD (2017) Systematic design for trait introgression projects. Theor Appl Genet 130:1993–2004

Canzar S, El-Kebir M (2011) A mathematical programming approach to marker-assisted gene pyramiding. In: Proceedings of the 11th workshop on algorithms in bioinformatics. Springer, pp 26–38

Castiglioni P, Ajmone-Marsan P, van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group ditsribution. Theor Appl Gen 99:425–431

CGC (2018) Crop germplasm committees. Briefings 2010–2018 USDA-ARS GRIN. https://www.ars-grin.gov/npgs/cgcweb.html

Chalyk ST (1994) Properties of maternal haploid maize plants and potential application to maize breeding. Euphytica 79:13–18

Char SN, Unger-Wallace E, Frame B, Briggs SA, Main M, Spalding MH, Vollbrecht E, Wang K, Yang B (2015) Heritable site-specific mutagenesis using TALENs in maize. Plant Biotechnol J 13:1002–1010

Char SN, Neelakandan AK, Nahampun H, Frame B, Main M, Spalding MH, Becraft PW, Meyers BC, Walbot V, Wang K, Yang B (2017) An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. Plant Biotechnol J 15:257–268

Chase SS (1949) Monoploid frequencies in a commercial double cross hybrid maize, and in its component single cross hybrids and inbred lines. Genetics 34:328–332

Chase SS (1951) Efficient methods of developing and improving inbred lines. The monoploid method of developing inbred lines. Report of 6th hybrid corn industry research conference, pp 29–34

Chase SS (1952) Production of homozygous diploids of maize from monoploids. Agron 44:263–267

Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132:1199–1210

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhajarvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807

Chilcoat D, Liu Z-B, Sander J (2017) Use of CRISPR/Cas9 for crop improvement in maize and soybean. Prog Mol Biol Transl Sci 149:27–46

Chojnacki S, Cowley A, Lee J, Foix A, Lopez R (2017) Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. Nucleic Acids Res 45:W550–W553

Chourey PS, Schwartz D (1971) Ethyl methanesulfonate-induced mutations of the Sh_1 protein in maize. Mutat Res 12:151–157

Ci X, Li M, Liang X, Xie Z, Zhang D, Li X, Lu Z, Ru G, Bai L, Xie C, Hao Z, Zhang S (2011) Genetic contribution to advanced yield for maize hybrids released from 1970 to 2000 in China. Crop Sci 51:13–20

Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 4:265–270

Coe EH Jr, Sarkar KR (1964) The detection of haploids in maize. Heredity 555:231–233

Coe EH, Sarkar KR (1966) Preparation of nucleic acids and a genetic transformation attempt in maize. Crop Sci 6:432–435

Coe E, Cone K, McMullen M, Chen SS, 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 128:9–12

Comstock RE, Robinson HF, Harvey PH (1949) A breeding procedure designed to make maximum use of both general and specific combining ability. Agron J 41:360–367

Cone KC, McMullen MD, Bi IV, Davis GL, Yim YS, Gardiner JM, Polacco ML, Sanchez-Villeda H, Fang Z, Schroeder SG, Havermann SA, Bowers JE, Paterson AH, Soderlund CA, Engler FW, Wing RA, Coe EH Jr (2002) Genetic, physical, and informatics resources for maize. On the road to an integrated map. Plant Physiol 130:1598–1605

Cooper M, Podlich DW (2002) The E(NK) model: extending the NK model to incorporate gene by environment interactions and epistasis for diploid genomes. Compexity 7:31–47

Cooper M, Podlich DW, Micallef KP, Smith OS, Jensen NM et al. (2002) Complexity, quantitative traits and plant breeding: a role for simulation modeling in the genetic improvement of crops. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CAB

Cooper M, Gho C, Leafgren R, Tang T, Messina C (2014) Breeding drought-tolerant maize hybrids for the US corn-belt: discovery to product. J Exp Bot 65:6191–6204

Cress CE (1967) Reciprocal recurrent selection and modifications in simulated populations. Crop Sci 7:561–567

Crow JF (1998) 90 years ago: the beginning of hybrid maize. Genetics 148:923–928

Crow JF (1999) Dominance and overdominance. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. ASA, CSSA, Madison, pp 49–58

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021–1031

Darrah DL, Zuber MS (1986) 1985 United States farm maize germplasm base and commercial breeding strategies. Crop Sci 26:1109–1113

Davenport CB (1908) Degeneration, albinism and inbreeding. Science 28:454–455

De Beukelaer H, De Meyer G, Fack V (2015) Heuristic exploitation of genetic structure in marker-assisted gene pyramiding problems. BMC Genet 16:2–16

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592–601

Dicke FF, Guthrie WD (1988) The most important corn insects. In: Sprague GF, Dudley JW (eds) Corn and corn improvement, 3rd edn. American Society of Agronomy, Madison, pp 767–868

Doebley J, Wendel JF, Smith JSC, Stuber CW, Goodman MM (1988) The origin of Cornbelt maize: the isozyme evidence. Econ Bot 42:120–131

Dollinger EJ (1954) Studies on induced mutation in maize. Genetics 39:750–766

Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Masignani V (2010) Structure and dynamics of the

pan-genome of *Streptococcus pneumoniae* and closely related species. Genome Biol 11:R107
Dong Q, Roy L, Freeling M, Walbot V, Brendel V (2003) ZmDB, an integrated database for maize genome research. Nucleic Acids Res 31:244–247
Dubreuil P, Dufour P, Krejci E, Causse M, deVienne D, Gallais A, Charcosset A (1996) Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. Crop Sci 36:790–799
Duvick DN (1965) Cytoplasmic pollen sterility in corn. Adv Genet 13:1–56
Duvick DN (1984) Genetic diversity in major farm crops on the farm and in reserve. Econ Bot 38:161–178
Duvick DN (2005a) Genetic progress in yield of United States maize (*Zea mays* L.). Maydica 50:193–202
Duvick DN (2005b) 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
East EM (1908) Inbreeding in corn. Rep Conn Agric Exp Stn 1907:419–428
Eberhart SA (1970) Factors affecting efficiencies of breeding methods. Afr Soils 15:669–680
Eder J, Chalyk ST (2002) In vivo haploid induction in maize. Theor Appl Genet 104:703–708
Edmeades GO, Trevisan W, Prasanna BM, Campos H (2017) Tropical maize (*Zea mays* L.). In: Campos H, Caligari PDS (eds) Genetic improvement of tropical crops. Springer, New York, pp 57–109
Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138
Einset J (1942) Chromosome length in relation to transmission frequency in maize trisomes. Genetics 28:349–364
Eisenstein M (2015) Startups use short-read data to expand long-read sequencing market. Nat Biotechnol 33:433–435
Emerson RA (1917) Genetical studies of variegated pericarp in maize. Genetics 2:1–35
Eynard SE, Croiseau P, Laloe D, Fritz S, Calus MPL, Restoux G (2018) Which individuals to choose to update the reference population? Minimizing the loss of genetic diversity in animal genomic selection programs. G3 8:113–121
FAOSTAT (2018) Crop data. FAO United Nations, Rome. http://www.fao.org/faostat/en/#data/QC
Fehr, WR (1991) Maximizing genetic improvement. In: Principles of cultivar development: theory and technique. Macmillian, USA, pp. 219–246
Feng L, Sebastian S, Smith S, Cooper M (2006) Temporal trends in SSR allele frequencies associated with long-term selection for yield of maize. Maydica 51:293–300
Feng PC, Qi Y, Chiu T, Stoecker MA, Schuster CL, Johnson SC, Fonseca AE, Huang J (2014) Improving hybrid seed production in corn with glyphosate-mediated male sterility. Pest Manag Sci 70:212–218
Fernandez J, Toro MA (1999) The use of mathematical programming to control inbreeding in selection schemes. J Anim Breed Genet 116:447–466
Fischer T, Byerlee D, Edmeades G (2014) Crop yields and global food security: will yield increase continue to feed the world? ACIAR monograph no. 158. Australian Centre for International Agricultural Research, Canberra, xxii + 634 pp
Fisher RA (1930) The fundamental theorem of natural selection. The genetical theory of natural selection. Oxford University Press, Oxford, pp 22–47
Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM (2009) Heterosis is prevalent for multiple traits in diverse maize germplasm. PLoS ONE 4:e7433
Frame BR, Shou H, Chikwamba RK, Zhang Z, Xiang C, Fonger TM, Pegg SE, Li B, Nettleton DS, Pei D, Wang K (2002) *Agrobacterium tumefaciens*-mediated transformation of maize embryos using a standard binary vector system. Plant Physiol 129:13–22
Fraser AS, Burnell DG (1970) Computer models in genetics. McGraw-Hill, San Franscisco
Frisch M, Bohn M, Melchinger AE (1999) Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci 39:1295–1301
Fromm ME, Taylor LP, Walbot V (1986) Stable transformation of maize after gene transfer by electroporation. Nature 319:791–793
Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578
Gabay-Laughnan S, Laughnan JR (1994) The male sterility and restorer genes in maize. In: Freeling M, Walbot V (eds) The maize handbook. Springer, New York, pp 418–423
Gaffney J, Anderson J, Franks C, Collinson S, MacRobert J, Woldemariam W, Albertsen MC (2016) Robust seed systems, emerging technologies and hybrid crops for Africa. Food Secur. 9:36–44
Gama EEG, Hallauer AR (1977) Relation between inbred and hybrid traits in maize. Crop Sci 17:703–706
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 6:28334
Gao C (2018) The future of CRISPR technologies in agriculture. Nat Rev Mol Cell Biol 19:275–276
Garcia AAF, Wang S, Melchinger AE, Zeng Z-B (2008) Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. Genetics 180:1707–1724
Gardiner JM, Coe EH, Melia-Hancock S, Hoisington DA, Chao S (1993) Development of a core RFLP map in maize using an immortalized F2 population. Genetics 134:917–930
Gardner CA (2012) Maize diversification by capturing useful alleles from exotic germplasm. In: Proceedings 48th Annual Illinois Corn Breeding School, March 5–6, 2012. Urbana-Champaign, IL, p 172
Garing F (2000) Inbred corn plant 90QDD1 and seeds thereof. United States Patent No. US 6,034,305. US Patent Office, Washington, DC
Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, Mackay IJ, Hickey JM (2017) A two-part strategy for using genomic selection to develop inbred lines. Crop Sci 57:2372–2386
Geiger HH (2009) Doubled haploids. Maize handbook—volume ii: genetics and genomics. Springer, New York, pp 641–657
Geiger HH, Braun MD, Gordillo GA, Koch S, Jesse J, Krutzfeldt BAE (2006) Variation for female fertility among haploid maize lines. Maize Genet Newsl 80:28–29
Georges F, Ray H (2017) Genome editing of crops: a renewed opportunity for food security. GM Crops & Food 8:1–12

Gibson PB, Brink RA, Stahmann MA (1950) The mutagenic action of mustard gas on *Zea mays*. J Hered 41:232–238

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 C, 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

Giraud H, Bauland C, Falque M, Madur D, Combes V, Jamin P, Monteil C, Laborde J, Palaffre C, Gaillard A, Blanchard P, Charcosset A, Moreau L (2017) Reciprocal genetics: identifying QTLs for general and specific combining abilities in hybrids between multiparental populations from two maize (*Zea mays* L.) heterotic groups. Genetics 207:1167–1180

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296:92–100

Golicz AA, Batley J, Edwards D (2016) Towards plant pangenomics. Plant Biotechnol J 14:1099–1105

Golovkin MV, Abraham M, Morocz S, Bottka S, Feder A, Dudits D (1993) Production of transgenic maize plants by direct DNA uptake into embryogenic proroplasts. Plant Sci 90:41–52

Gonzalez VH, Tollenaar M, Bowman A, Good B, Lee EA (2018) Maize yield potential and density tolerance. Crop Sci 58:472–485

Goodman MM (1978) A brief survey of the races of maize and current attempts to infer racial relationships. In: Walden DB (ed) Maize breeding and genetics, pp143–184

Goodman MM (1999) Broadening the genetic diversity in maize breeding by use of exotic germplasm. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops, pp139–148

Goodman MM (2005) Broadening the U.S. maize germplasm base. Maydica 50:203–214

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186

Gordillo GA, Geiger HH (2008) Optimization of DH-line based recurrent selection procedures in maize under a restricted annual loss of genetic variance. Euphytica 161:141–154

Gordon-Kamm WJ, Spencer TM, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams WR Jr, Willets NG, Rice TB, Mackey CJ, Krueger RW, Kausch AP, Lemaux PG (1990) Transformation of maize cells and regeneration of fertile transgenic plants. Plant Cell 2:603–618

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

Gorjanc G, Gaynor RC, Hickey JM (2018) Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor Appl Genet 131:1953–1966

Gowen JW (1952) Heterosis. Iowa State College Press, Ames

Graham GI, Wolff DW, Stuber CW (1997) Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. Crop Sci 37:1601

Grimsley N, Hohn T, Davies JW, Hohn B (1987) *Agrobacterium* mediated delivery of infectious maize streak virus into maize plants. Nature 325:177–179

Gurian-Sherman D (2009) Failure to yield: evaluating the performance of genetically engineered crops. Union of Concerned Scientists. http://www.ucsusa.org/assets/documents/food_and_agriculture/failure–to–yield.pdf

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

Hallauer AR, Miranda F (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames

Hallauer AR, M. J. Carena, Filho JBM (2010) Selection: experimental results. In: Quantitative genetics in maize breeding. Handbook of plant breeding, vol 6. Springer, New York, pp 291–383

Han Y, Cameron JN, Wang L, Beavis WD (2017) The predicted cross value for genetic introgression of multiple alleles. Genetics 205:1409–1423

Häntzschel KR, Weber G (2010) Blockage of mitosis in maize root tips using colchicine-alternatives. Protoplasma 241:99–104

Hazel LN (1943) The genetic basis for constructing selection indices. Genetics 28:476–490

Heady EO (1954) Simplified presentation and logical aspects of linear programming technique. J Farm Econ 36:1035–1048

Heather JM, Chain B (2016) The sequence of sequencers: the history of sequencing DNA. Genomics 107:1–8

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection of crop improvement. Crop Sci 49:1–12

Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Gen 72:761–769

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447

Herzog E, Frisch M (2011) Selection strategies for marker-assisted backcrossing with high-throughput marker systems. Theor Appl Genet 123:251–260

Herzog E, Falke KC, Presterl T, Scheuermann D, Ouzunova M, Frisch M (2014) Selection strategies for the development of maize introgression populations. PLoS ONE 9:e92429

Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52:146–152

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Hillel J, Schaap T, Haberfeld A, Jeffreys AJ, Plotzky Y, Cahaner A, Lavi U (1990) DNA fingerprints applied to gene introgression in breeding programs. Genetics 124:783–789

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Penagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N, Kaeppler SM, Buell CR (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26:121–135

Holland JB (2004) Breeding: incorporation of exotic germplasm. In: Goodman RM (ed) Encyclopedia of plant and crop science. Marcel Dekker, New York, pp 222–224

Holland J, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 22:9–112

Hospital F (2001) Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. Genetics 158:1363–1379

Hospital F, Charcosset A (1997) Marker-assisted introgression of quantitative trait loci. Genetics 147:1469–1485

Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132:1199–1210

Howard R, Carriquiry AL, Beavis WD (2014) Parametric and non-parametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. G3 (Bethesda) 4:1027–1046

Howard JT, Pryce JE, Baes C, Maltecca C (2017) Invited review: inbreeding in the genomics era: inbreeding, inbreeding depression, and management of genomic variability. J Dairy Sci 100:6009–6024

Huang CR, Burns KH, Boeke JD (2012) Active transposition in genomes. Annu Rev Genet 46:651–675

Hufford MB, Lubinksy P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibara J (2013) Correction: the genomic signature of crop-wild introgression in maize. PLOS Genetics. https://doi.org/10.1371/annotation/2eef7b5b-29b2-412f-8472-8fd7f9bd65ab

Hull RH (1945) Recurrent selection and specific combining ability in corn. J Am Soc Agron 37:134–145

Inghelandt DV, 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

Initiative AG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

ISAAA (2017) Global status of commercialized Biotech/GM Crops in 2017: biotech crop adoption surges as economic benefits accumulate in 22 years. ISAAA Brief no. 53, ISAAA: Ithaca, NY

Ishida Y, Saito H, Ohta SH, Hiei Y, Komari T, Kumashiro T (1996) High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. Nat Biotechnol 14:745–750

Ishii T, Yonezawa K (2007) Optimization of the marker-based procedures for pyramiding genes from multiple donor lines: I. Schedule of crossing between the donor lines. Crop Sci 47:537–547

Jannink J-L (2010) Dynamics of long-term genomic selection. Genet Sel Evol 42:11

Jeffrey B, Lübberstedt T (2014) Molecular breeding of bioenergy traits. In: Corn S, Goldman (ed.) Compendium of bioenergy plantsscience. Publishers/Taylor & Francis/CRC PRESS, Boca Raton, FL, USA, pp.198–215

Jenkins MT (1940) The segregation of genes affecting yield of grain in maize. J Am Soc Agron 32:55–63

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44:812–815

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, 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

Johnson I, Eldredge J (1953) Performance of recovered popcorn inbred lines derived from outcrosses to dent corn. Agron J 45:105–110

Johnson B, Gardner CO, Wrede KC (1988) Application of an optimization model to multi-trait selection programs. Crop Sci 28:723–728

Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. Genetics 2:466–479

Jugenheimer RJ (1985) Corn improvement, seed production and uses. RE Krieger, Malabar, p 794

Kadam DC, Potts SM, Bohn MO, Lipka AE, Lorenz AJ (2016) Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. G3(6):3443–3453

Kaeppler S (2012) Heterosis: many genes, many mechanisms—end the search for an undiscovered unifying theory. ISRN Bot 2012:1–12

Karush W (1939) Minima of functions of several variables with inequalities as side constraints. University of Chicago, Chicago

Kassie GT, Erenstein O, Mwangi W, La Rovere R, Setimela P, Langyintuo A (2012) Characterization of maize production in southern Africa: synthesis of CIMMYT/DTMA household level farming system surveys in Angola, Malawi, Mozambique, Zambia and Zimbabwe. Socio-economics program working paper 4. CIMMYT, Mexico, D.F

Kato A (2002) Chromosome doubling of haploid maize seedling using nitrous oxide gas at the flower primordial stage. Plant Breed 1215:370–377

Kelliher T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio ML, Green J, Chen Z, McCuiston J, Wang W, Liebler T, Bullock P, Martin B (2017) MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature 542:105–109

Kermicle JL (1969) Androgenesis conditioned by a mutation in maize. Science 166:1422–1424

Kermicle JL (1994) Indeterminate gametophyte ig biology and use. In: Freeling M, Walbot V (eds) The maize handbook. Springer, New York, pp 388–393

Kinghorn BP (1998) Mate selection by groups. J Dairy Sci 81:55–63

Kingsbury N (2009) Hybrid: the history and science of plant breeding. The University of Chicago Press, Chicago

Kleiber D, Prigge V, Melchinger AE, Burkard F, San Vicente F, Palomino G, Gordillo GA (2012) Haploid fertility in temperate and tropical maize germplasm. Crop Sci 52:623–630

Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WI, Menz MA, Franks CD, Klein PE (2008) The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci 48:12

Klein RR, Miller FR, Bean S, Klein PE (2016) Registration of 40 converted germplasm sources from the reinstated sorghum conversion program. J Plant Regist 10:57

Kremling KAG, Chen S-Y, Su M-H, Lepak NK, Romay MC, Swarts KL, Lu F, Lorant A, Bradbury PJ, Buckler ES (2018) Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature 555:520–523

Kuhn HW, Tucker AW (1951) Nonlinear programming. In: Proceedings of 2nd Berkeley symposium, pp 481–492

Kumar GR, Sakthivel K, Sundaram RM, Neeraja CN, Balachandran S, Rani NS, Viraktamath B, Madhav M (2010) Allele mining in crops: prospects and potentials. Biotechnol Adv 28:451–461

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–168

Laborda PR, Oliveira KM, Garcia AF, Paterniani MEAG, Souza AP (2005) Tropical maize germplasm: what can we say about its genetic diversity in the light of molecular markers? Theor Appl Genet 111:1288–1299

Lanza LLB, de Souza CL Jr, Ottoboni LMM, Vieira MLC, de Souza AP (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor Appl Genet 94:1023–1030

Larkins JR (2000) Inbred corn plant RQAA8 and seeds thereof. U.S. Patent No 6,143,961. US Patent Office, Washington DC

Lawrence CJ, Harper LC, Schaeffer ML, Sen TZ, Seigfried TE, Campbell DA (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. Int J Plant Genom 2008:496957

Le Clerc V, Bazante F, Baril C, Guiard J, Zhang D (2005) Assessing temporal changes in genetic diversity of maize varieties using microsatellite markers. Theor Appl Genet 110:294–302

Leakey ADB, Uribelarrea M, Ainsworth EA, Naidu SLO, Rogers A, Ort DR, Long SP (2006) Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO_2 concentration in the absence of drought. Plant Physiol 140:779–790

Lee M, Phillips RL (1987) Genomic rearrangements in maize induced by tissue culture. Genome 29:123–128

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 48:453–461

Leung H, Raghavan C, Zhou B, Oliva R, Choi IR, Lacorte V, Jubay ML, Cruz CV, Gregorio G, Singh RK (2015) Allele mining and enhanced genetic recombination for rice breeding. Rice 8:1

Li Y, Ma X, Wang T, Li Y, Liu C, Liu Z, Sun B, Shi Y, Song Y, Carlone M, Bubeck D, Bhardwaj H, Whitaker D, Wilson W, Jones E, Wright K, Sun S, Niebur W, Smith S (2011) Increasing maize productivity in China by planting hybrids with germplasm that responds favorably to higher planting densities. Crop Sci 51:2391–2400

Li X, Zhu C, Wang J, Yu J (2012a) Computer simulation in plant breeding. Adv Agron 116:219–264

Li X, Zhu C, Yeh CT, Wu W, Takacs EM, Petsch KA, Tian F, Bai G, Buckler ES, Muehlbauer GJ, Timmermans MC, Scanlon MJ, Schnable PS, Yu J (2012b) Genic and nongenic contributions to natural variation of quantitative traits in maize. Genome Res 22:2436–2444

Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS, Zuo Q, Shi XH, Li YF, Zhang WK, Hu Y, Kong G, Hong HL, Tan B, Song J, Liu ZX, Wang Y, Ruan H, Yeung CK, Liu J, Wang H, Zhang LJ, Guan RX, Wang KJ, Li WB, Chen SY, Chang RZ, Jiang Z, Jackson SA, Li R, Qiu LJ (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol 32:1045–1052

Li R, Hsieh CL, Young A, Zhang Z, Ren X, Zhao Z (2015) Illumina synthetic long read sequencing allows recovery of missing sequences even in the "Finished" C. elegans Genome. Sci Rep 5:10814

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 (2016) Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. Plant J 86:391–402

Li H, Rasheed A, Hickey LT, He Z (2018) Fast-forwarding genetic gain. Trends Plant Sci 23:184–186

Liang Z, Zhang K, Chen K, Gao C (2014) Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. J Genet Genom 41:63–68

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Liu F, Zhu Y, Yi Y, Lu N, Zhu B, Hu Y (2014) Comparative genomic analysis of *Acinetobacter baumannii* clinical isolates reveals extensive genomic variation and diverse antibiotic resistance determinants. BMC Genom 15:1163

Liu Z, Ren J, Trampe B, Frei UK, Lübberstedt T (2016) Doubled haploids: from obscure phenomenon to key technology of current maize breeding programs. Plant Breed Rev 40:123–166

Liu C, Li X, Meng D, Zhong Y, Chen C, Dong X, Xu X, Chen B, Li W, Li L, Tian X, Zhao H, Song W, Luo H, Zhang Q, Lai J, Jin W, Yan J, Chen S (2017) A 4-bp insertion at *ZmPLA1* encoding a putative phospholipase A generates haploid induction in maize. Mol Plant 10:520–522

Longin CFH, Utz HF, Reif JC, Wegenast T, Schipprack W, Melchinger AE (2007) Hybrid maize breeding with doubled haploids: III. Efficiency of early testing prior to doubled haploid production in two-stage selection for tescross performance. Theor Appl Genet 115:519–527

Longin CFH, Mi X, Wurschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor Appl Genet 128:1297–1306

Lowe K, Wu E, Wang N, Hoerster G, Hastings C, Cho MJ, Scelonge C, Lenderts B, Chamberlin M, Cushatt J, Wang L, Ryan L, Khan T, Chow-Yiu J, Hua W, Yu M, Banh J, Bao Z, Brink K, Igo E, Rudrappa B, Shamseer PM, Bruce W, Newman L, Shen B, Zheng P, Bidney D, Falco C, Register J, Zhao ZY, Xu D, Jones T, Gordon-Kamm W (2016) Morphogenic regulators Baby boom and Wuschel improve monocot transformation. Plant Cell 28:1998–2015

Lu Y, Yan J, Guimaraes CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Parentoni SN, Shah T, Rong T, Crouch JH, Xu Y (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120:93–115

Lu Y, Shah T, Hao Z, Taba S, Zhang S, Gao S, Liu J, Cao M, Wang J, Bhanu Pralash A, Rong TXuY (2011) Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. PLoS ONE 6(9):e24861

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

Magorokosho C (2006) Genetic diversity and performance of maize varieties from Zimbabwe, Zambia and Malawi. PhD thesis Texas A&M University, College Station, TX, 179 pp

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

Mangelsdorf PC (1961) Introgression in maize. Euphytica 10:157–168

Mariani C, De Beuckeleer M, Truettner J, Leemans J, Goldberg RB (1990) Induction of male sterility in plants by a chimaeric ribonuclease gene. Nature 347:737–741

Markelz RJ, Strellner RS, Leakey ADB (2011) Impairment of C4 photosynthesis by drought is exacerbated by limiting nitrogen and ameliorated by elevated CO_2 in maize. J Exp Bot 62:3235–3246

Marulanda JJ, Mi X, Melchinger AE, Xu JL, Wurschum T, Longin CF (2016) Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. Theor Appl Genet 129:1901–1913

Mastrodomenico AT, Hendrix CC, Below FE (2018) Nitrogen use efficiency and the genetic variation of maize expired plant variety protection germplasm. Agric Agric 8:3

Masuka B, Atlin GN, Olsen M, Magorokosho C, Labuschagne M, Crossa J, Banziger M, Pixley KV, Vivek B, Biljon A, MacRobert JF, Alvarado G, Prasanna BM, Makumbi D, Makumbi D, Tarekegne AT, Das B, Zaman-Allah M, Cairns JE (2017a) Gains in maize genetic improvement in Eastern and Southern Africa : I. CIMMYT hybrid breeding pipeline. Crop Sci 57:168–179

Masuka B, Magorokosho C, Olsen M, Atlin GN, Bänziger M, Pixley KV, Vivek BS, Labuschagne M, Matemba-Mutasa R, Burgueño J, Macrobert J, Prasanna BM, Das B, Makumbi D, Tarekegne A, Crossa J, Zaman-Allah M, van Biljon A, Cairns JE (2017b) Gains in maize genetic improvement in Eastern and Southern Africa: II. CIMMYT open-pollinated variety breeding pipeline. Crop Sci 57:180–191

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E, Doebley J (2002) A single domestication for maize shown by

multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084
May BP, Liu H, Vollbrecht E, Senior L, Rabinowicz PD, Roh D, Pan X, Stein L, Freeling M, Alexander D, Martienssen R (2003) Maize-targeted mutagenesis: a knockout resource for maize. Proc Natl Acad Sci USA 100:11541–11546
McCarty DR, Suzuki M, Hunter C, Collins J, Avigne WT, Koch KE (2013) Genetic and molecular analyses of UniformMu transposon insertion lines. Methods Mol Biol 1057:157–166
McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36:344–355
MCGC (2016) Maize crop germplasm committee. USDA-ARS GRIN. Vulnerability statement recommendations. https://www.ars-grin.gov/npgs/cgc_reports/maizevuln2016.pdf. Accessed 12 Dec 2016
Melchinger AE, Geiger HH, Schnell FW (1986) Epistasis in maize (*Zea mays* L.). Theor Appl Genet 72:231–239
Melchinger AE, Schipprack W, Mi X, Mirdita V (2015) Oil content is superior to oil mass for identification of haploid seeds in maize produced with high-oil inducers. Crop Sci 55:188–195
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 USA 106:21019–21026
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
Mi X, Utz HF, Technow F, Melchinger AE (2014) Optimizing resource allocation for multistage selection in plant breeding with R package. Crop Sci 54:1413
Mikel MA, Dudley JW (2006) Evolution of North American dent corn from public to proprietary germplasm. Crop Sci 46:1193–1205
Mir C, Zerjal T, Combes V, Dumas F, Madur D, Bedoya C, Dreisigacker S, Franco J, Grudloyma P, Hao P, Hearne S, Jampatong C, Laloë D, Muthamia Z, Nguyen T, Prasanna B, Taba S, Xie C, Yunus M, Zhang S, Warburton M, Charcosset A (2013) Out of America: tracing the genetic footprints of the global diffusion of maize. Theor Appl Genet 126:2671–2682
National Corn Growers Association (2018) World corn production, National Corn Growers Association (sourced from USDA, FAS Grain: World Markets and Trade) http://www.worldofcorn.com/#world-corn-production. Accessed 12 Jan 2018
Nelson PT, Goodman MM (2008) Evaluation of elite exotic maize inbreds for use in temperate breeding. Crop Sci 48:85–92
Nelson PT, Krakowsky MD, Coles ND, Holland JB, Bubeck DM, Smith JSC, Goodman MM (2016) Genetic characterization of the North Carolina State University maize lines. Crop Sci 56:259–275
Neuffer MG (1957) Additional evidence on the effect of X-ray and ultraviolet radiation on mutation in maize. Genetics 42:273–282
Neuffer MG (1994) Mutagenesis. In: Freeling M, Walbot V (eds) The maize handbook. Springer, New York, pp 212–218
Neuffer MG, Coe EH (1978) Paraffin oil technique for treating mature corn pollen with chemical mutagens. Maydica 23:21–28
Neuffer MG, Fiscor G (1963) Mutagenic action of ethyl methanesulfonate in maize. Science 139:1296–1297
Neuffer MG, Johal G, Chang MT, Hake S (2009) Mutagenesis—the key to genetic analysis. In: Bennetzen JL, Hake S (eds) The maize handbook. Springer, New York, pp 63–84
Niu X, Xie R, Liu X, Zhang F, Li S, Gao S (2013) Maize yield gains in Northeast China in the last six decades. J Integr Agric 12:630–637
NRC (1972) Committee on genetic vulnerability of major crops. (1972) Genetic vulnerability of major crops. Natl Acad Sci Washington DC, 307 pp
NRC (1993) Committee on managing global genetic resources: agricultural imperatives. Board on agriculture. Natl Res Council National Academy Press, Washington DC
Pace J, Gardner C, Romay C, Ganapathsybrumanian B, Lübberstedt T (2015) Genome-wide association analysis of seedling root development in maize. BMC Genom 16:47
Paten B, Novak AM, Eizenga JM, Garrison E (2017) Genome graphs and the evolution of genome inference. Genome Res 27:665–676
Peccoud J, Velden KV, Podlich D, Winkler C, Arthur L, Cooper M (2004) The selective values of alleles in a molecular network model are context dependent. Genetics 166:1715–1725
Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CA, McMullen MD, Holland JB, Bradbury PJ, Buckler ES (2014) The genetic architecture of maize height. Genetics 196:1337–1356
Peng T, Sun X, Mumm RH (2014a) Optimized breeding strategies for multiple trait integration: I Minimizing linkage drag in single event introgression. Mol Breed 33:89–104
Peng T, Sun C, Mumm RH (2014b) Optimized breeding strategies for multiple trait integration: II Process efficiency in event pyramiding and trait fixation. Mol Breed 33:105–115
Peterson P (1953) A mutable pale green locus in maize. Genetics 38:682–683
Piepho H-P (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49:1165–1176
Piperno DR, Ranere AJ, Holst I, Inarte 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 USA 106:5019–5024
Pixley KV (2006) Hybrid and open-pollinated varieties in modern agriculture. In: Lamkey KR, Lee M (eds) Plant breeding: the Arnel R. Hallauer international symposium. Blackwell Publishing, Ames
Podlich DW, Cooper M (1998) Qu-GENE: a simulation platform for quantitative analysis of genetic models. Bioinformatics 14:632–653
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 108:6893–6898
Pollacsek M (1992) Management of the ig gene for haploid induction in maize. Agronomie 12:247–251
Pong-Wong R, Woolliams JA (2007) Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. Genet Sel Evol 39:3–25
Prasanna BM (2012) Diversity in global maize germplasm: characterization and utilization. J Biosci 37:843–855
Puchta H, Hohn B (2010) Breaking news: plants mutate right on target. Proc Natl Acad Sci USA 107:1165–11658
Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res 26:342–350
Qin X, Feng F, Li Y, Xu S, Siddique KHM, Liao Y (2016) Maize yield improvements in China: past trends and future directions. Plant Breed 135:166–176
Randolph LF (1940) Note on haploid frequencies. Maize Genet Coop Newsl 14:23–24
Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA (2012) Recent patterns of crop yield growth and stagnation. Nat Commun 3:1293
Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, Beck S, Bohn M, Frisch M (2003) Use of SSRs for establishing heterotic groups in subtropical maize. Theor Appl Genet 107:947–957

Reif JC, Fischer S, Schrag TA, Lamkey KR, Klein D, Dhillon BS, Utz HF, Melchinger AE (2010) Broadening the genetic base of European maize heterotic pools with US Cornbelt germplasm using field and molecular marker data. Theor Appl Genet 120:301–310

Ren J, Wu P, Tian X, Lübberstedt T, Chen SJ (2017) Fine mapping of quantitative trait locus qhmf4 causing haploid male fertility in maize based on segregation distortion. Theor Appl Genet 130:1349–1359

Rendel JM, Robertson A (1950) Estimation of gnetic gain in milk yield by selection ina closed herd of dairy cattle. Journal of Genetics 50:1–8

Rhoades M (1931) Cytoplasmic inheritance of male sterility in Zea mays. Science 73:340–341

Rhoades MM (1938) Effect of *Dt* gene on the mutability of the *a1* allele in maize. Genetics 23:377–397

Rhodes CA, Pierce DA, Mettler IJ, Mascarenhas D, Detmer JJ (1988) Genetically transformed maize plants from protoplasts. Science 240:204–207

Robertson A (1957) Optimum group size in progeny testing and family selection. Biometrics 13:442–450

Robertson A (1960) A theory of limits in artificial selection. Proc R Soc Lond 153:234–249

Rogers DL, McGuire PE (2015) Genetic erosion: context is key. In: Ahuja MR, Jain SM (eds) Genetic diversity and erosion in plants. Springer, New York, pp 1–24

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

Romero Navarro JA, Willcox M, RomayC Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, OrtegaA Banda AE, Montiel NO, Ortiz-Monasterio I, Vicente FS, EspinozaAG Atlin G, WenzlP Hearne S, Buckler S (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet 49:476–480

Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. Science 281(363):365

Rotarenco VA, Dicu G, State D, Fuia S (2010) New inducers of maternal haploids in maize. Maize Genet Coop Newslett 84:1–7

Sanchez D, Liu S, Ibrahim R, Blanco M, Lübberstedt T (2018) Association mapping of seedling root traits in exotic derived doubled haploid lines of maize. Plant Sci 268:30–38

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sarvella P, Grogan CO (1967) The mutagenic effects of gamma rays on *Zea mays* in relation to ear location. Radiat Bot 7:107–111

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE 6:e17855

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schneerman MC, Charbonneau M, Weber DF (2000) A survey of ig containing materials. Maize Genet Coop Newslett 74:92–93

Schrag TA, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, Melchinger AE (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. Genetics 208:1373–1385

Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK (1993) Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. Science 262:110–114

Segerman B (2012) The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. Front Cell Infect Microbiol 2:116

Servin B, Martin OC, Mezard M, Hospital F (2004) Toward a theory of marker-assisted gene pyriamiding. Genetics 168:513–523

Shi J, Gao H, Wang H, Lafitte R, Archibald RL, Yang M, Hakimi SH, Mo H, Habben J (2017) ARGOS8 variants generated by CRISPR-Cas9 improve maize grain under field drought stress conditions. Plant Biotechnol J 15:2017–2216

Shukla VP, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM, Rock JM, Wu YY, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD, Urnov FD (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459:437–441

Shull GH (1908) The composition of a field of maize. Am Breeders Assoc Rep 4:296–301

Singleton WR (1941) Hybrid vigor and its utilization in sweet corn breeding. Am Nat 75:48–60

Smith HF (1936) A discriminant function for plant selections. Ann Eugenetics 7:240–250

Smith OS (1986) Covariance between line per se and testcross performance. Crop Sci 26:540–543

Smith JSC, Smith OS (1991) Restriction fragment length polymorphisms can differentiate among U.S. maize hybrids. Crop Sci 31:893–899

Smith DR, White DG (1988) Diseases of corn. In: Sprague GF, Dudley JW (eds) Corn and corn improvement, III edn. American Society of Agronomy, Madison, pp 687–766

Smith JSC, Smith OS, Wright S, Wall SJ, Walton W (1992) Diversity of U.S. hybrid maize germplasm as revealed by restriction fragment length polymorphisms. Crop Sci 32:598–604

Smith S, Cooper M, Gogerty J, Löffler C, Borcherding D, Wright K (2014) Maize. In: Smith et al (ed) Yield gains in major U.S. field crops. CSSA Spec. Publ. 33. ASA, CSSA, and SSSA, Madison, pp 125–171

Smith JS, Gardner CA, Costich DE (2017) Ensuring the genetic diversity of maize and its wild relatives. In: Watson D (ed) Achieving sustainable cultivation of maize. Burleigh Dodds, Cambridge

Springer NM, Stupar RM (2007) Allelic variation and heterosis in maize: how do two halves make more than a whole? Genome Res 17:264–275

Springer N, Anderson SN, Andorf C, Ahern K, Bai F, Barad O, Barbazuk WB, Bass HW, Baruch K, Ben-Zvi G, Buckler ES, Bukowski R, Campbell MS, Cannon EKS, Chomet P, Dawe RK, Davenport R, Dooner HK, Du LH, Du C, Easterling KA, Gault C, Guan J-C, Jander G, Hunter CT, Jiao Y, Koch KE, Kol G, Kudo T, Li Q, Lu F, Mayfield-Jones D, Mei W, McCarty DR, Noshay J, Ronen G, Settles MA, Shem-Tov D, Shi J, Soifer I, Stein JC, Suzuki M, Vera DL, Vollbrecht E, Vrebalov JT, Ware D, Wei X, Wimalanathan K, Woodhouse MR, Xiong W, Brutnell TP (2018) The W22 genome: a foundation for maize functional genomics and transposon biology. Nat Genet 50(9):1282–1288

St Martin SA, Skavaril RV (1984) Computer simulation as a tool in teaching introductory plant breeding. J Agron Educ 13:43–47

Stadler LJ (1949) A note on haploidy in maize (unpublished)

Stadler LJ, Sprague GF (1936) Genetic effects of ultra-violet radition in maize. II. Filtered raditions. Genetics 22:579–583

Stadler LJ, Uber F (1942) Genetic effects of ultra-violet radiation in maize.IV. Comparison of monochromatic radiations. Genetics 27:84–118

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Sun X, Peng T, Mumm RH (2011) The role and basics of computer simulation in support of critical decision in plant breeding. Mol Breed 28:421–436

Sun C, Hu Z, Zheng T, Lu K, Zhao Y, Wang W, Shi J, Wang C, Lu J, Zhang D, Li Z, Wei C (2017) RPAN: rice pan-genome browser for approximately 3000 rice genomes. Nucleic Acids Res 45:597–605

Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, Liu H, Ma X, Jiao Y, Wang B, Wei X, Stein JC, Glaubitz JC, Lu F, Yu G, Liang C, Fengler K, Li B, Rafalski A, Schnable PS, Ware DH, Buckler ES, Lai J (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nat Genet 50:1289–1295

Svitashev S, Young JK, Schwartz C, Gao H, Falco SC, Cigan MA (2015) Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 guide RNA. Plant Physiol 169:931–945

Svitashev S, Schwartz C, Lenderts B, Young JK, Cigan MA (2016) Genome editing in maize by CRISPR-Cas9 ribonucleoprotein complexes. Nat Commun 7:13274

Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peeples M, Lepak N, Prim L, Cinta Romay M, Ross-Ibarra J, de Jesus Sanchez-Gonzalez J, 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 357:512–515

Technow F, Messina CD, Totir LR, Cooper M (2015) Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS ONE 10:e0130855

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

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

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci USA 102:13950–13955

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 43:159–162

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biol 4:12

Tinker NA, Mather DE (1993) GREGOR: software for genetic simulation. J Hered 84:237

Troyer AF (1999) Background of U.S. hybrid corn. Crop Sci 39:601–626

Troyer AF (2006) Adaptedness and heterosis in corn and mule hybrids. Crop Sci 46:528–543

Troyer AF, Wellin EJ (2009) Heterosis decreasing in hybrids: yield test inbreds. Crop Sci 49:1969–1976

Unterseer S, Pophaly SD, Peis R, Westermeier P, Manfred M, Seidel MA, Haberer G, Mayer KFX, Ordas B, Pausch H, Tellier A, Bauer E, Schon C-C (2016) A comprehensive study of the genomic differentiation between temperate Dent and Flint maize. Genome Biol 17:137

Unterseer S, Seidel MA, Bauer E, Haberer G, Hochholdinger F, Opitz N, Marcon C, Baruch K, Spannagl M, Mayer KFX, Schön C-C (2017) European Flint reference sequences complement the maize pan-genome. bioRxiv https://doi.org/10.1101/103747

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 USA 108:1088–1092

van Heerwaarden J, Hufford MB, Ross-Ibarra J (2012) Historical genomics of North American maize. Proc Natl Acad Sci USA 109:12420–12425

Vanous A, Gardner C, Blanco M, Martin-Schwarze A, Flint-Garcia S, Bohn M, Edwards J, Lübberstedt T (2018) Association mapping of flowering and plant height traits in germplasm enhancement of maize doubled haploid (GEM-DH) lines. The Plant Genome 11:170083

Vernikos G, Medini D, Riley DR, Tettelin H (2015) Ten years of pan-genome analyses. Curr Opin Microbiol 23:148–154

Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sánchez GJ, Doebley J (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95:1240–1253

Visscher PM, Haley CS, Thompson R (1996) Marker-assisted introgression in backcross breeding programs. Genetics 144:1923–1932

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55:641–658

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, Muszynski M, Brendel VP, Brutnell TP (2010) Genome-wide distribution of transposed Dissociation elements in maize. Plant Cell 22:1667–1685

Voss-Fels K, Snowdon RJ (2016) Understanding and utilizing crop genome diversity via high-resolution genotyping. Plant Biotechnol J 14:1086–1094

Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genetics 10:e1004845

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

Wang AS, Evans RA, Altendorf PR, Hanten JA, Doyle MC, Rosichan JL (2000) A mannose selection system for production of fertile transgenic maize plants from protoplasts. Plant Cell Rep 19:654–660

Wang K, Frame B, Marcell L (2003a) Maize genetic transformation. In: Jaiwal PK, Singh RP (eds) Plant genetic engineering, vol 2. Improvement of food crops. Sci-Tech Publication, Houston, pp 175–217

Wang X, Van Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, DeLacy IH, Cooper M, Rajaram S (2003b) Comparison of two breeding strategies by computer simulation. Crop Sci 43:1764–1773

Wang J, van Ginkel M, Trethowan R, Ye G, DeLacy I, PodlichD Cooper M (2004) Simulating the effects of dominance and epistasis on selection response in the CIMMYT wheat breeding program using QuCim. Crop Sci 44:2006–2018

Wang J, Chapman SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47:582–590

Watson A, Ghosh S, Williams M, Cuddy WS, Simmonds J, Rey M-D, Hatta MAM, Hinchlife A, Steed A, Reynolds D, Adamski N, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan M, Edwards D, Batley J, Raman H, Rogers C, Domoney C, Moore G, Harwood W, Nicholson P, Dieters MJ, DeLacy IH, Zhou J, Uauy C, Boden SA, Park RF, Wulf BBH, Hickey LT (2017) Speed breeding: a powerful tool to accelerate crop research and breeding. Nat Plants 4:23–29

Weber D, Helentjaris T (1989) Mapping RFLP loci in maize using B–A translocations. Genetics 121:583–590

Wei F, Zhang J, Zhou S, He R, Schaeffer M, Collura K, Kudrna D, Faga BP, Wissotski M, Golser W, Rock SM, Graves TA, Fulton RS, Coe E, Schnable PS, Schwartz DC, Ware D, Clifton SW, Wilson RK, Wing RA (2009) The physical and genetic framework of the maize B73 genome. PLoS Genet 5:e1000715

Wellhausen EJ, Roberts LM, Hernandez X, Mangelsdorf PC (1952) Races of maize in Mexico: their origin, characteristics and distribution. Bussey Inst Harvard Univ Cambridge, Mass, p 222

Wen W, Araus JL, Shah T, Cairns J, Mahuku G, Bänziger M, Torres JL, Sánchez C, Yan J (2011) Molecular characterization of a diverse maize inbred line collection and its potential utilization for stress tolerance improvement. Crop Sci 51:2569–2581

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(10):e47832

Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. Genet Res 75:249–252

Wilcox JR, Cavins JF (1995) Backcrossing high seed protein to a soybean cultivar. Crop Sci 35:1036–1041

Williams ME (2016) Alternative mutagens for maize (*Zea mays* L.). Maize Genom Genet 7:1–8

Winston WL, VenkataramananM, Goldberg JB (2003) Introduction to mathematical programming, vol 1. Operations Research, 4 edn. Brooks/Cole, Pacific Grove, CA

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 8:e1000409

Woolliams JA, Berg P, Dagnachew BS, Meuwissen TH (2015) Genetic contributions and their optimization. J Anim Breed Genet 132:89–99

Wu Y, Frei UK, Liu H, De La Fuente G, Huang K, Wei Y, Lübberstedt T (2015) Combining genomic selection and doubled haploid technology increases efficiency of maize breeding. In: Govil JN (ed) Recent developments in biotechnology, vol 2. Plant Biotechnology. Studium Press, pp 215–237

Wu Y, Fox TW, Trimnell MR, Wang L, Xu RJ, Cigan AM, Huffman GA et al (2016) Development of a novel recessive genetic male sterility system for hybrid seed production in maize and other cross-pollinating crops. Plant Biotechnol J 14:1046–1054

Wych RD (1988) Production of hybrid seed corn. In: Sprague GF (ed) Corn and corn improvement. American Society of Agronomy Inc, Crop Science Society of America, and Soil Science Society of America, Madison, pp 565–607

Xing HL, Dong L, Wang ZP, Zhang HY, Han CY, Liu B, Wang XC, Chen QJ (2014) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol 14:327

Xu Y (2016) Envirotyping for deciphering environmental impacts on crop plants. Theor Appl Genet 129:653–673

Xu P, Wang L, Beavis WD (2011) An optimization approach to gene stacking. Eur J Oper Res 214:168–178

Xu Y, Li P, Zou C, Lu Y, Xie C, Zhang X, Prasanna BM, Olsen MS (2017) Enhancing genetic gain in the era of molecular breeding. J Exp Bot 68:2641–2666

Yang N, Xu X-W, Wang R-R, Peng W-L, Cai L, Song J-M, Li W, Luo X, Niu L, Wang Y, Jin M, Chen L, Luo J, Deng M, Wang L, Pan Q, Liu F, Jackson D, Yang X, Chen L-L, Yan J (2017a) Contributions of *Zea mays* subspecies mexicana haplotypes to modern maize. Nat Commun 8:1874

Yang J, Mezmouk S, Baumgarten A, Buckler ES, Guill KE, McMullen MD, Mumm RH, Ross-Ibarra J (2017b) Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. PLoS Genet 13:e1007019

Ye G, Smith KF (2008) Marker-assisted gene pyramiding for inbred line development: basic principles and practical guidelines. Int J Plant Breed 2:1–10

Yim YS, Davis GL, Duru NA, Musket TA, Linton EW, Messing JW, McMullen MD, Soderlund CA, Polacco ML, Gardiner JM, Coe EH Jr (2002) Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiol 130:1686–1696

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296:79–92

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–555

Yu X, Li X, Guo T, Zhu C, Wu Y, Mitchell SE, Roozeboom KL, Wang D, Wang ML, Pederson GA, Tesso TT, Schnable PS, Bernardo R, Yu J (2016) Genomic prediction contributing to a promising global strategy to turbocharge gene banks. Nat Plants 2:16150

Zabirova ER, Shatskaya OA, Shcherbak VS (1993) Line 613/2 as a source of a high frequency of spontaneous diploidization in corn. Maize Genet Coop Newsl 67:67

Zhang X, Zhang H, Li L, Lan H, Ren Z, Liu D, Wu L, Liu H, Jaqueth J, Li B, Pan G, Gao S (2016) Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. BMC Genom 17:697

Zhang D, Wu S, An X, Xie K, Dong Z, Zhou Y, Xu L, Fang W, Liu S, Liu S, Zhu T, Li J, Rao L, Zhao J, Wan X (2018) Male-sterile line and hybrid seed production based on the ZmMs7 gene encoding a PHD-finger transcription factor. Plant Biotech J 16:459–471

Zhao Z-Y, Gu W, Cai T, Tagliani L, Hondred D, Bond D, Schroeder S, Rudert M, Pierce D (2001) High throughput genetic transformation mediated by *Agrobacterium tumefaciens* in maize. Mol Breed 8:323–333

Zhao Y, Mette MF, Reif JC (2015) Genomic selection in hybrid breeding. Plant Breed 134:1–10

Zila CT, Ogut F, Romay MC, Gardner CA, Buckler ES, Holland JB (2014) Genome-wide association study of Fusarium ear rot disease in the. BMC Plant Biol 14:372

Zuber MS, Darrah DL (1981) 1979 U.S. corn germplasm base. In: Proceedings of the 35th Ann Corn and Sorghum Ind Res Conf. ,Washington DC American Seed Trade Association, pp 234–249

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.