ABSTRACT
Commercial-scale plant breeding is a complex process in which new crop varieties are continuously being developed to improve yield and agronomic performance over current varieties. A wide array of naturally occurring genetic changes are sources of new characteristics available to plant breeders. During conventional plant breeding, genetic material is exchanged that has the potential to beneficially or adversely affect plant characteristics. For this reason, commercial-scale breeders have implemented extensive plant selection practices to identify the top-performing candidates with the desired characteristics while minimizing the advancement of unintended changes. Selection practices in maize (Zea mays L.) breeding involve phenotypic assessments of thousands of candidate lines throughout hundreds of different environmental conditions over many years. Desirable characteristics can also be introduced through genetic modification. For genetically modified (GM) crops, molecular analysis is used to select transformed plants with a single copy of an intact DNA insert and without disruption of endogenous genes. All the while, GM crops go through the same extensive phenotypic characterization as conventionally bred crops. Data from both conventional and GM maize breeding programs are presented to show the similarities between these two processes.
characteristics (yield, quality, resistance to abiotic and biotic stresses, etc.) and devising strategies to combine these characteristics to obtain superior varieties (Acquaah, 2012). In its simplest form, plant breeding results in improved crop varieties (the commercial product, also referred to as cultivars or hybrids, depending on the crop) through the mating of two or more parental lines that contain desirable characteristics. The target characteristics are measured over multiple generations throughout different environments and stress conditions. Offspring with desirable characteristics are selected, whereas offspring with undesired characteristics are eliminated from further breeding. The degree of improvement in the new variety depends on the level of genetic variation affecting the characteristics of interest and the ability to accurately measure the expression of these characteristics in many different environmental conditions (Fehr et al., 1987).

Genetic engineering, commonly referred to as genetic modification, is an additional tool that affords plant breeders new sources of characteristics, such as genes that confer abiotic or biotic stress tolerance, with many of these genes not available in the crop’s genome (Prigge and Melchinger, 2012; Weber et al., 2012; Prado et al., 2014; Schnell et al., 2015). After a genetically modified (GM) line containing the desired DNA insert is chosen, the DNA insert is introduced (via backcrossing) into well-characterized, conventionally bred elite varieties. The selection process that follows is essentially the same as is used for conventionally bred crops.

Using hybrid maize (Zea mays L.) as the example, data are presented from Monsanto case studies to illustrate the commercial-scale breeding practices used to supply seed to farmers. The range of sources of genetic variation, extent of testing, and scope of plant selection processes used for conventional breeding are presented first, followed by a parallel overview for GM varieties.

**CONVENTIONAL PLANT BREEDING**

**Sources of Genetic Variation Used for Conventional Breeding Programs**

Plant breeders improve crops by identifying sources of genetic variation for the characteristics of interest. Plant genomes (the genetic material in each species) are highly variable, even within and among closely related species (Weber et al., 2012). Table 1 shows various natural biological processes that create genetic diversity. These include the movement of transposable elements, vertical gene flow via mating with wild relatives, horizontal gene flow (Bock, 2010) from unrelated plants, Agrobacterium, florenroviruses, pararetroviruses, and mutations such as single-nucleotide polymorphisms, chromosomal rearrangements, and the presence, absence, or copy number of germline genes.

Maize, in particular, has a high level of sequence and structural diversity (Buckler et al., 2006; Springer et al., 2009; Lai et al., 2010). A genomic comparison of two maize inbreds, B73 and Mo17, revealed an unprecedented level of genomic structural diversity compared with most higher eukaryotes studied thus far (Springer et al., 2009). For example, by a conservative estimate, at least 180 putative single-copy genes were present in one inbred but absent in the other, and >400 instances of putative sequence copy number variation between the two inbreds were observed (Springer et al., 2009). Likewise, a comparison (Hirsch et al., 2016) between B73 and PH207 found >2500 genes present only in one of those inbreds.

The traditional perspective has been that conventional breeding does not introduce new genes, only variations (alleles) of already existing genes. However, the emergence of the pangenome concept (Golicz et al., 2016) makes it clear that conventional breeding results in the introduction of additional genes and alleles, as well as novel combinations of genes. It is evident that the same mechanisms of genome instability found in nature that lead to genetic diversity (Table 1) are also active during conventional breeding (Weber et al., 2012; Schnell et al., 2015). One example is from a recent comparison of DNA structure (both large chromosomal changes and single-nucleotide polymorphisms) across a collection of soybean [Glycine max (L.) Merr.] cultivars, many derived by conventional breeding (Anderson et al., 2016). This study showed that genetic changes accumulated spontaneously across many conventional germplasm (i.e., standing variation). Another recent study of many maize varieties (both conventional and GM) showed that most of the observed compositional differences were associated with the backcrossing practices from conventional breeding (Venkatesh et al., 2015). Repetitive DNA sequences and structural variations in plants have the potential to contribute to genetic change. Similarly, transposable genetic elements in maize and many other plant species can mediate genetic changes (Hirsch and Springer, 2017). Transposable elements are DNA sequences that can change position within a genome, resulting in small insertions and deletions, as well as larger rearrangements such as inversions, deletions, and duplication of genes (Zhang and Peterson, 2004; Zhang et al., 2006; Weber et al., 2012).

Horizontal gene transfer across phylogenetic boundaries is another natural process that results in genetic variation in plants (Bock, 2010; Soucy et al., 2015), including transfer of DNA from bacteria, viruses, and unrelated plants (Berghthorsson et al., 2003; Staginnus et al., 2007; Liu et al., 2012; El Bairdouri et al., 2014; Geering et al., 2014; Kyndt et al., 2015). One recent example was a study demonstrating that a wide selection of sweet potato [Ipomoea batatas (L.) Lam.] varieties contain Agrobacterium transfer DNA with expressed genes (Kyndt et
its region of origin and can also be used as a source of novel characteristics for breeding in other world regions. However, differences in agronomic characteristics, such as photoperiod sensitivity, temperature response, and disease susceptibility, must be recognized when using germplasm from other parts of the world.

To gain additional diversity not present in the existing germplasm, random genetic mutations can be induced using chemical or radiation mutagenesis (IOM/NRC, 2004; Curry, 2016). Breeders then select for agriculturally desirable genetic changes while selecting against the many unintended or unwanted changes that can occur with these methods (Bolon et al., 2011). Over 3000 plant varieties, mostly vegetables, fruits, grains, and ornamentals (IAEA-MVD, 2017), have been developed via mutagenesis. Well-known examples include the Star Ruby grapefruit (Citrus ´ paradisi Macfad) (Hensz, 1971) and high-oleic canola (Brassica napus L.) oil (Auld et al., 2015) as a consequence of natural transformation of an ancestral form of the plant.

Breeders commonly use locally adapted, domesticated germplasm (often called landraces) that exhibit exceptional performance in a specific group of geographic or management conditions, as well as international germplasm that are adapted to and have been selected for a wide range of environmental conditions (Hallauer et al., 1988; Lynch and Walsh, 1998; Doebley, 2004; Acquaah, 2012; Butruille et al., 2015). Locally adapted germplasm contain agronomic characteristics required for high yields in their environment and meet regional consumer preferences; therefore, they may be used more easily within a breeding program for their geographically distinct region.

Seed developers often have germplasm collections for crops that include most of the major cultivation regions of the world (Butruille et al., 2015; Smykal et al., 2015). This germplasm is adapted for breeding in

<table>
<thead>
<tr>
<th>Genetic change</th>
<th>Genotypic or phenotypic example</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transposable elements (transposons)</td>
<td>White grapes, blood oranges</td>
<td>Lisch (2013)</td>
</tr>
<tr>
<td></td>
<td>&gt;25,000 unique insertions detected across 31 varieties of soybean</td>
<td>Tian et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Yellow maize</td>
<td>Palaisa et al. (2003)</td>
</tr>
<tr>
<td></td>
<td>&gt;50 new insertions of a transposon per rice plant per generation</td>
<td>Naito et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Elongated tomato fruit</td>
<td>Xiao et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>Round or wrinkled peas (Mendel)</td>
<td>Ellis et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>2 million transposons exchanged between higher plants</td>
<td>El Baldouri et al. (2014)</td>
</tr>
<tr>
<td>Organellar DNA in nuclear DNA</td>
<td>Gain and loss of mitochondrial DNA common to maize inbred lines</td>
<td>Lough et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>Gain and loss of chloroplast DNA common to maize inbred lines</td>
<td>Roark et al. (2010)</td>
</tr>
<tr>
<td>Bacterial genes</td>
<td>Expression of several bacterial genes in sweet potatoes</td>
<td>Kyndt et al. (2015)</td>
</tr>
<tr>
<td>Crossing with wild relatives</td>
<td>&gt;60 wild relatives have been used for &gt;100 characteristics (80% involve pest or disease resistance) in 13 crops</td>
<td>Hajjar and Hodgkin (2007)</td>
</tr>
<tr>
<td></td>
<td>Dozens of alien genes used in wheat breeding</td>
<td>Jones et al. (1995)</td>
</tr>
<tr>
<td>Pararetroviruses</td>
<td>Stable viral DNA in rice genome</td>
<td>Liu et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Stable viral DNA in tomato (previously also seen in potato)</td>
<td>Stagginnus et al. (2007)</td>
</tr>
<tr>
<td>Fiorendoviruses</td>
<td>Stable integrations in all plants</td>
<td>Geering et al. (2014)</td>
</tr>
<tr>
<td>Insertions and deletions</td>
<td>Submergence-tolerant rice</td>
<td>Xu et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Dwarf sorghum</td>
<td>Multani et al. (2003)</td>
</tr>
<tr>
<td></td>
<td>Yellow soybean seeds</td>
<td>Tuteja et al. (2004)</td>
</tr>
<tr>
<td>Single-nucleotide polymorphisms (SNPs)</td>
<td>Maize proteins (300–400 amino acids long) from 2 alleles differ by 3–4 amino acids</td>
<td>Tenallon et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Maize genome has 55 million SNPs</td>
<td>Gore et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Green Revolution gene has 2 SNPs for dwarf wheat</td>
<td>Peng et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>One SNP caused loss of shattering in domestic rice</td>
<td>Konishi et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Tall or short pea plants (Mendel)</td>
<td>Ellis et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>7 new SNPs created per meiosis per billion base pairs</td>
<td>Ossowski et al. (2010)</td>
</tr>
<tr>
<td>Presence, absence, or copy number of genes</td>
<td>856 wild-type soybean genes absent in cultivated varieties (and &gt;186,000 DNA insertions or deletions)</td>
<td>Lam et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>&gt;10^7 SNPs, 30,000 insertion or deletions, and a few large chromosomal deletions (&gt;18 genes) in 6 elite maize varieties</td>
<td>Lai et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Copy number variation relates to soybean cyst nematode resistance</td>
<td>Cook et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Pinot Noir, Corvina, and Tannat wine grapes have 1873 genes not found in other wine grapes</td>
<td>Da Silva et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Only 81% of Brassica genes are always present in the same number</td>
<td>Golicz et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>2500 genes found only in either B73 or PH207</td>
<td>Hirsch et al. (2016)</td>
</tr>
<tr>
<td></td>
<td>G. soja genotypes can vary by 1000 to 3000 gene families from each other</td>
<td>Li et al. (2014)</td>
</tr>
</tbody>
</table>
Nevertheless, breeders are shifting from using these random methods of introducing genetic diversity over the past couple of decades to newer, more predictable methods like genetic engineering. Gene editing is emerging as a new tool for plant breeders to introduce alterations in genes to achieve desired plant characteristics; however, due to the scope of this review, gene editing is not discussed in detail here.

**Interaction of Genotype and Environment on Plant Characteristics**

A plant variety’s appearance and performance (phenotype, $P$) is determined by an interaction between its genes (genotype, $G$) and the environment ($E$), commonly expressed as $P = G \times E$ (Lynch and Walsh, 1998). Traditionally, a major task of the plant breeder has been to differentiate between the effects of environment and genotype. The experimental design and selection strategy that breeders use to identify the most desirable genetic material is determined by the heritability, environment, and correlations between characteristics. Qualitative characteristics result in a limited number of possible phenotypes that can be placed into defined categories. For example, in simple Mendelian inheritance, only white or purple flower color, long or short stems, and yellow or green pod color are possible in pea (*Pisum sativum* L.) and are the result of a single gene controlling each characteristic (Ellis et al., 2011). Qualitative characteristics are more reliably expressed across different environments and are said to have high heritability. By comparison, quantitative characteristics are expressed as a continuous variation of phenotypes, in which few to multiple genes and the environment contribute to the expression of the characteristic. Examples of complex, quantitative characteristics in maize include yield and plant height, which are more strongly influenced by the environment than simpler qualitative characteristics (Hallauer et al., 1988).

**Unintended Effects in Conventional Plant Breeding**

Genomes are dynamic, and changes have occurred throughout evolution and during breeding that have the potential to affect plant characteristics. However, unforeseen hazardous effects (e.g., the production of an allergen, toxin, or other hazard that was not previously known to exist for the plant species) have not been documented (Parrott et al., 2012; Steiner et al., 2013; Ladics et al., 2015a, 2015b). Many crops have at least one characteristic known to be a hazard in some scenarios (IOM/NRC, 2004). For instance, potato (*Solanum tuberosum* L.) contains glycoalkaloids, such as solanine, and legumes contain lectins that are toxic when consumed in high doses or uncooked. Changes in the amounts of solanine in potato have occurred during breeding (Smith, 2013). Known anti-nutrients and toxins like these are typically monitored during conventional breeding to ensure that their levels are not increased (Zywicki et al., 2005; Shepherd et al., 2015). Increases in known toxicants, while unintended, are foreseeable and thus manageable in a breeding program.

Segregation of undesirable alleles already present in the breeding population also leads to unintended or undesirable effects. An example of a potentially undesirable agronomic characteristic is the degree to which the husk covers the ear in maize, which affects the protection against moisture and disease provided by the husk and in turn affects the geographic range of adaptation for maize hybrids. Maize grown in the southeastern United States require good husk coverage to prevent ear rot, whereas maize grown in northern latitudes benefit from reduced husk coverage to allow for faster ear dry down. Therefore, breeders developing short-season varieties (northern germplasm) from longer-season varieties (southern germplasm) must actively select against long, tight husks.

Other types of unintended effects from plant breeding may only be observed in certain environmental conditions. For example, a condition known as virescence (a temporary light-green color observed in maize seedlings) is triggered by cold and can result in delays in maturity (Fig. 1) (Hopkins and Walden, 1977). This phenotype is commonly observed in maize varieties that originated in warm climates, like those found in parts of South America. Plant breeding programs using South American maize varieties may contain this unintended virescence characteristic in their new varieties unless the allele responsible for this
phenotype is selected against. Table 2 shows how adjusting the size of the breeding population and the intensity of selection can help to maintain genetic gain when a simply inherited characteristic, such as virescence, must be selected against in a breeding program. Typically, when an undesirable phenotype is present in the population at a high frequency, a plant breeding program will scale up the size of the beginning population to improve the chances of finding superior candidates with the desired characteristics that also lack the undesirable phenotype. It is important to note that unintended phenotypic changes, for the most part, do not equate with a food or feed safety risk.

**BREEDING AND SELECTION OF INBRED LINES AND HYBRIDS**

Breeding hybrid crops involves the creation of both the inbred lines, which may have improved characteristics important for hybrid seed production, and hybrids, which are the final commercial products that contain the combined characteristics from crossing two inbred lines. Hybrids usually show dramatic improvements over the inbred lines due to heterosis (Hallauer et al., 1988; Doebley, 2004; Schnable et al., 2009; Springer et al., 2009; Acquaah, 2012). Some crops, such as maize, have well-identified heterotic pools, making it easier to know which inbred line combinations tend to give superior hybrids. In other hybrid crops, such as tomato (*Solanum lycopersicum* L.) and cucumber (*Cucumis sativus* L.), there is less understanding of heterotic pools, and thus less predictability of which inbred line combinations will produce superior hybrids.

The selection process for inbreds and hybrids is very similar, with the hybrid development process illustrated in Fig. 2. In both cases, the process starts with large populations of candidate lines that are evaluated for numerous agronomic characteristics in a small number of locations. Each field season, the top-performing candidate lines within a breeding population are advanced into the next field season’s testing. As materials advance, a reduced number of candidate populations are tested in an increasing number of locations and with larger plots at each location. This procedure often means that, when summed, hundreds of thousands of individual plants are grown in hundreds of different environments. This ensures that the potential commercial hybrids are tested in numerous environments.

Table 2. Impact of inheritance of virescence on the size of the starting population and the selected number of individuals advanced from an *F*₂ to an *F*₃ generation. Plant breeders can adjust selection intensity and population size to account for undesirable characteristics when inheritance for the characteristic is predictable. The undesirable phenotype in this example is represented by the single-gene expression of virescence, and the desired phenotype is homozygous recessive.

<table>
<thead>
<tr>
<th>Breeding strategy</th>
<th>Starting <em>F</em>₂ population size</th>
<th>Phenotypic selection intensity</th>
<th>Intensity of marker-assisted selection of nonvirescence</th>
<th>Individuals advanced to <em>F</em>₃ generation</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virescence locus homozygous (no segregation for virescence allele)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard selection practice</td>
<td>2000</td>
<td>12</td>
<td>N/A</td>
<td>240</td>
<td>Adequate no. of individuals and genetic gain target</td>
</tr>
<tr>
<td>Virescence locus heterozygous (segregation for virescence allele)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No adjustment in selection intensity</td>
<td>2000</td>
<td>12</td>
<td>25</td>
<td>60</td>
<td>Inadequate no. of individuals to reach pipeline target</td>
</tr>
<tr>
<td>Relax selection intensity</td>
<td>2000</td>
<td>48</td>
<td>25</td>
<td>240</td>
<td>Reduced genetic gain</td>
</tr>
<tr>
<td>Increase population size</td>
<td>8000</td>
<td>12</td>
<td>25</td>
<td>240</td>
<td>Adequate no. of individuals and genetic gain target</td>
</tr>
</tbody>
</table>
and stressors to understand the durability of the genetics and to monitor for unintended agronomic or performance effects that might only be observed in certain conditions.

In developing a new maize inbred, a breeder crosses two parental lines to produce a new segregating population that will be evaluated for numerous agronomic characteristics (Fig. 3). If one or both of these parental lines are elite (and possibly have been used in other commercial breeding programs), the offspring in this breeding population may not exhibit significant unintended or inferior characteristics but will simply need to be evaluated to identify the best performing amongst them. Sometimes a non-elite or nonadapted parent may need to be used to introduce the desired characteristics. In this case, a specific combination of parents can unexpectedly combine inferior genetics for a characteristic (e.g., delayed maturity, inadequate husk coverage, susceptibility to disease, virulence), and many of the individuals within the population will be discarded as unsuitable for further breeding due to the deficiency. As described below, marker-assisted selection can be applied to enable very specific selection of the genes of the elite germplasm with desirable characteristics while selecting against the undesired genes from the non-elite parent (Butruille et al., 2015, Chang and Coe, 2009, Eathington et al., 2007).

The top-performing candidates are repeatedly crossed back to the elite parent (backcrossing eliminates 50% of the non-elite genome with each cross), followed by ongoing evaluation and selection of the offspring with the most desired agronomic characteristics while removing any offspring with undesired characteristics. By the sixth backcross, the selected offspring will theoretically contain >99% of the DNA from the elite parent and <1% from the non-elite parent (Fig. 4). The offspring developed through this process should display the characteristics of interest from the non-elite parent, few or none of the undesirable characteristics from the non-elite parent, and all of the desired characteristics of the elite parent.

**TECHNOLOGIES USED TO EXPEDITE THE BREEDING PROCESS**

Creating inbred lines through self-pollination is a time-consuming process, as numerous generations of plants must be grown to maturity. Doubled-haploid technology is used to produce homozygous materials more quickly (Chang and Coe, 2009; Prigge and Melchinger, 2012). Doubled-haploid methods start from an initial population

---

**Fig. 3. Agronomic characteristics observed during selection of inbreds and hybrids in conventional breeding.** The agronomic characteristics listed are observed at very specific growth stages during the plant lifecycle. These agronomic and phenotypic measurements provide data for maize breeders to select varieties for advancement in breeding programs. 1 Visual appraisal of the vegetative plant on a 1-to-9 scale: 1 = excellent vigor to 9 = poor vigor; 2 Days from planting until ~50% of the plants are showing the characteristic.
of plants in which haploid variants (containing a single set of chromosomes) are developed and then treated to induce duplication of their chromosomes to produce diploid plants with identical chromosomal pairs. This new inbred is a genetically stable line that can undergo observation and testing to determine its value as a breeding parent.

During the generation of inbred lines, marker-assisted selection can be applied to enrich for characteristics of interest and the favorable quantitative trait loci (QTLs) that underlie the characteristics (Eathington et al., 2007). For instance, Monsanto’s breeding program uses a subset of >50 different QTLs that exist for each of the characteristics of days to silking, days to anthesis, and grain oil content (Laurie et al., 2004; Buckler et al., 2009). Additionally, there are >100 QTLs associated with the development of root architecture in maize (Zurek et al., 2015), and breeding companies screen for a targeted subset of them within their breeding pipeline. Markers associated with virescence and other undesirable characteristics are used to identify, and subsequently discard, progeny that contain these undesired characteristics while advancing plants with the desirable QTLs. Seed chipping technology, combined with marker-assisted selection, allows plant breeders to select against undesired alleles with greater efficiency and accuracy, and at a reduced cost (Butruille et al., 2015). Seed chipping technology allows for identification of the genetics of the candidate lines from a small tissue sample taken from the seed. This negates the need to grow all the seeds in the field to identify the best lines.

**SELECTION FOR ADAPTABILITY TO VARIABLE ENVIRONMENTAL CONDITIONS**

**Geographic Adaptation**

The specific requirements for a given crop can differ greatly by the geographic location in which the crop is being grown. For example, some crop diseases can be found in many parts of the world, such as northern corn leaf blight [**Exserohilum turcicum** (Pass.) K.J. Leonard & E.G. Suggs] (Leonard et al., 1989), and resistance to these diseases is an important characteristic in most maize hybrids. Other diseases are specific to certain areas. Commercial-scale plant breeding programs must constantly evaluate the different geographic needs for each crop when determining which characteristics to select for in their breeding program (Butruille et al., 2015). For example, in the western US plains (e.g., Nebraska and Kansas), maize products with exceptional tolerance to Goss’s wilt [**Clavibacter michiganensis** subsp. **nebraskensis** (Vidaver & Mandel) Davis et al.] are required by growers. The presence of the pathogen, environmental conditions (specifically the frequency of hail events), and management practices (such as continuous maize rotations) increase the importance of tolerance to this disease when developing maize hybrids for the western plains. Table 3 shows the shift in disease screening efforts to develop commercial products with improved Goss’s wilt tolerance. A fivefold increase in early-generation screening for Goss’s wilt resulted in approximately five times more commercial products with acceptable tolerance to the disease.

**Environmental Effects and the Need for Testing across Multiple Seasons**

Repeatability is an important factor when considering the commercial potential of new genetics. Temperature, water availability, solar radiation, and insect and disease pressures all vary by environment and by season, along with agronomic practices. Multienvironment and multiyear trials allow breeders to test their candidates for variations in performance among different geographies, year-to-year environmental variations, and response to different types of management...
Table 3. Representation of a shift in screening efforts in the Monsanto North America corn breeding program with enhanced screening in 2008 to improve tolerance to Goss’s wilt.†

<table>
<thead>
<tr>
<th>First-year testing</th>
<th>Commercial deployment year</th>
<th>Early pipeline families screened</th>
<th>Available screening locations</th>
<th>Corn products with moderate resistance or greater %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>2011</td>
<td>9,800</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>2008</td>
<td>2014</td>
<td>49,000</td>
<td>6</td>
<td>61</td>
</tr>
</tbody>
</table>

† Based on Monsanto internal data and seed portfolio ratings.

(Butruille et al., 2015). For example, most farmers will plant products with a relative maturity (R.M) that matches the historic R.M average of their region, such as 105 R.M. However, given weather fluctuations from year to year, the measure of growing degree units accumulation that represents a 105 R.M zone actually occurs <40% of the time.

Testing and selection across multiple seasons is critical to ensure that products are well adapted to future weather conditions and fluctuations. Compared with US Midwest agricultural climatic conditions seen during 1971 to 2000, the projected climatic conditions in 2070 to 2099 for many important weather variables (e.g., the number of consecutive days with rain or the number of days with frost or high nighttime temperatures) will change by >30 d, requiring new varieties suited for those new weather conditions (Hatfield et al., 2014).

**BREEDING OF GM CROPS**

**Sources of Candidate Genes for GM Crops and Early Testing**

The use of transgenes has become a high-profile complement to conventional breeding. Conventional plant breeding has successfully improved crops through selection practices that capture genetic gains, even though information on the specific genes and genetic networks that contribute to the desired agronomic characteristics is usually limited. In contrast, genetic engineering requires prior knowledge of the desired gene(s) for introduction into the plant. Several recent articles provide an overview of the GM crop development process from discovery through commercialization (Privalle et al., 2012; Mumm, 2013; Prado et al., 2014; NAS, 2016). The first step in the development of a GM crop is to identify a gene that confers the desired characteristic (referred to as a “trait” in GM crops). For this publication, genetic modification and GM crops refer to plants with a particular gene (or multiple genes) isolated from its source and directly introduced into the plant’s genome, and the end product is a new variety containing the desired trait (Ricroch and Hénard-Damave, 2016). Once inserted, the gene responsible for the desired trait is inherited by conventional Mendelian genetics during subsequent breeding (Weber et al., 2012). In most cases, the copied gene expresses a protein that confers the desired trait in the crop (Prigge and Melchinger, 2012; Prado et al., 2014). This review will focus on genes that encode proteins that confer the desired trait, using an herbicide resistance trait as an example.

The sources of genes range from the crop itself (and its close relatives) to more distantly related plant species or microbes, but not animal sources. Prior to making any GM crops, all potential proteins that would be produced by the inserted gene (called a “transgene”) are screened using computer algorithms to ensure that they are not similar to known allergens, toxins, antinutritional proteins, disease agents, or pharmacologically active proteins (Delaney et al., 2008; Silvanovich et al., 2009; Hammond et al., 2013). For example, comparing candidate proteins to known allergens is typically done using sequence alignment programs, such as BLAST or FASTA, and by using a so-called sliding eight-amino-acid window search. Due to the precautionary nature of the sliding window search, ~15% of candidate sequences for transformation will yield a hit with proteins in an allergen database (Silvanovich et al., 2006). Any genes with hits with proteins in the allergen database are removed from further development.

During discovery, a large number of gene products are identified and tested by in vitro assays and/or in noncrop plants [such as Arabidopsis thaliana (L.) Heynh.], and ultimately within the intended crop (Mumm, 2013). The desired gene is cloned into a plasmid vector that can be transferred into a plant cell. A DNA construct (called an expression cassette) includes the gene(s) of interest and additional DNA sequences needed for gene expression in the plant. This expression cassette is part of the vector, as is a selectable marker that enables the identification of successfully transformed cells and progeny plants. In recent years, the trend is to include multiple genes of interest within a single DNA construct (Weber et al., 2012; NAS, 2016). For example, for effective insect control, different mechanisms of action can be identified by a combination of competitive binding experiments and testing new control agents against insects resistant to other insecticides (Pardo-López et al., 2013; Ladics et al., 2015a, 2015b; Chakroun et al., 2016; Jerga A et al., 2016; Jurat-Fuentes and Crickmore, 2017; Moar et al., 2017). Typically, dozens of versions of an expression cassette are tested in plants, both in greenhouse and confined field trials, to determine which best optimizes the gene expression needed to ensure commercial viability and reproducible efficacy of the trait. Furthermore, when
the cassettes contain multiple genes (allowing more biotech traits to be expressed within a single inserted genetic locus), even more rigorous testing is conducted to ensure that all of the genes are expressed as expected. This process of gene discovery and experiments to confirm that the gene achieves the desired characteristic (often called “proof of concept” studies) can take 2 to 3 yr (Privalle et al., 2012; Mumm, 2013; Prado et al., 2014). For single-gene qualitative traits (e.g., herbicide tolerance with a well-understood plant biochemical pathway), a large percentage of candidate genes will advance out of this proof-of-concept stage.

**BASICS OF AGROBACTERIUM-MEDIATED DNA INSERTION**

Once the gene of interest in the expression cassette has passed the proof of concept studies, a transformation vector consisting of the transgene and a selectable marker is designed and used to transform cultured plant cells (Prado et al., 2014). *Agrobacterium tumefaciens* Smith & Townsend is a soil bacterium that naturally inserts pieces of DNA into relatively random spots within plant genomes (Mehrotra and Goyal, 2012; Bourras et al., 2015). *Agrobacterium* is unique among plant pathogens in that it causes disease (called “crown gall”) by transferring genes into plant cells that produce compounds that aid further infection. These disease-causing *Agrobacterium* genes can be replaced with desirable genes (Mehrotra and Goyal, 2012; Bourras et al., 2015). This feature of *Agrobacterium* is exploited to insert the transgene and selectable marker into cultured plant cells. The plant cells are then screened on selection medium to identify transformed plant cells that can be regenerated into transgenic events (Schnell et al., 2015). The term “event” is commonly used to refer to each unique clone produced from DNA transformation. The random nature of the insertion into the plant’s genome results in events with different molecular characteristics and expression profiles; therefore, additional screening is needed to select the most desirable event, a process that is similar to how germplasm variation is used to select the desired phenotype in conventional breeding.

**THE EVENT SELECTION PROCESS**

The goal of event selection is to use multiyear, multi-generational testing to identify and remove events that lack the desired trait for the product concept, and those that demonstrate undesired characteristics (Privalle et al., 2012; Mumm, 2013; Prado et al., 2014). Data from a recently completed event selection process for the development of a new herbicide-tolerant GM maize hybrid is used as a specific example.

**SCREENING THE INITIAL GENERATION OF TRANSGENIC PLANTS**

In the example shown in Fig. 5, transformation resulted in the generation of 5236 individual, unique transgenic events (the initial generation [R0]) that each contained a two-transgene insert and a selectable marker gene. During the initial screening process, high-throughput polymerase chain reaction (PCR) assays were performed on leaf samples to identify and eliminate events that contained either more than one inserted copy of the desired DNA and/or contained vector backbone DNA (part of the DNA used by the bacterium for the DNA insertion process but not part of the expression cassette). In this example, 3936 events were eliminated and 1300 events were advanced (Fig. 5).

Next, the remaining 1300 events were analyzed for gene expression and effectiveness (trait efficacy). Cultured cells were treated with the herbicide, and events were selected that showed tolerance to the herbicide. In addition, events were further screened for commercial-level herbicide tolerance by applying the target herbicide and propagating the tolerant plants. After the R0 trait efficacy evaluation, 642 events were advanced (Fig. 5). If the transgene had been for above-ground insect control, leaf bioassays could be used for early screening to select for efficacy against the target insect. For other characteristics that are not as easily selectable, gene expression is measured at this stage, and events that do not express the target proteins and genes at a determined threshold would be discarded.

Depending on the product concept and configuration of the transformation vector, it may be desirable to remove the selectable marker from the commercial product, in which case a linkage assay is run for each event to determine whether the insertion of the selectable marker is independent of the transgene. Linkage means that the transgene and selectable marker are inserted close together on a chromosome and are unlikely to segregate away from each other in the offspring; therefore, they would almost
always be inherited together. In unlinked events, the marker is inserted far enough away from the transgene insertion that it will segregate during the next generation (R1) and can be eliminated (Matthews et al., 2001; Miller et al., 2002). In this example, removal of the selectable marker was desired, and events with linked selectable markers were discarded; 85 events had unlinked markers and were advanced (Fig. 5).

Additional molecular screening that is more comprehensive than the high-throughput PCR screening described above was applied in the R0 stage to ensure: (i) that only one copy of the DNA insert was in the selected events, (ii) the intactness and integrity of the DNA insert, (iii) the absence of any undesired DNA from the transformation process (e.g., the vector backbone), and (iv) that the DNA insertion had not interrupted endogenous genes (Kovalic et al., 2012). In this example, 54 events were advanced after in-depth R0 molecular analysis (Fig. 5). The discarded events contained either additional copies, fragments, or rearranged copies of the inserted DNA, or the inserted DNA interrupted endogenous genes. In addition to characterization of the DNA insertion, the RNA and/or proteins resulting from the inserted DNA were characterized to confirm that the intended gene products were being produced.

While R0 stage molecular screening and trait efficacy was progressing, general plant health was also monitored, and events with poor health and/or unintended phenotypes (“off-types”) were discarded. Examples of off-types include plants that show phenotypes such as poor germination, bleached tissues, discoloration, reduced plant height, or delays in silking or flowering. For this example, 5182 of the initial 5236 events (99%) were removed from the commercial product development pipeline prior to the completion of the initial generation (R0). The preliminary data for the remaining 54 R0 events showed that each had a single, intact, correct copy of the DNA insert, the insert had not disrupted any known endogenous genes, and the early field-testing results had shown acceptable levels of trait efficacy and no obvious “off-type” phenotypic characteristics.

**FIELD SCREENING TO SELECT LEAD EVENT**

The remaining 54 R0 events were self-pollinated (to generate R1 inbred seed) and outcrossed with elite lines (to generate F1, hybrid seed). The resulting seeds were advanced to small field trial evaluations for continued measurement of trait efficacy and numerous agronomic characteristics (Fig. 5 and 6). Similar to the process illustrated in Fig. 2 for conventional breeding, as the number of GM events in the population pool decreased, the number of field trial locations and replications for the remaining events increased, with additional testing occurring at each stage (e.g., phenotypic, trait efficacy). In this example, 22 and then 20 events were advanced through small-scale R1 and R2 generation field screens, respectively (Fig. 5 and 6). This was followed by first seven, then five, and eventually only two of the initial 5236 events being extensively evaluated in four successive large-scale field trials (Fig. 5 and 6), with the outcome being selection of a single highly suitable event for advancement to commercial development. By the time the single event was selected for commercialization and entered both regulatory safety evaluations and trait introgression (the next stages of commercial development, discussed later, Fig. 6), >300,000 individual plants were grown and observed in >20 inbred or hybrid lines (Table 4) throughout many different environmental and stress conditions.

In addition to the event selection process, a rigorous seed quality process is in place to ensure that the seed used for regulatory safety studies and trait introgression is from the selected event and has not cross-pollinated with other events (Fig. 6). In maize, this process typically takes two generations and begins by tracing the selected event back to a single ear produced from a single homozygous plant. In the example detailed above, molecular assays were performed to confirm the identity of the plant producing this single ear and the resulting progeny, and plants whose identity could not be confirmed were discarded. In addition to the confirmation of identity, PCR analyses were performed for all other events grown within 200 m that might have shed pollen within a 3-wk time period of the pollination of the commercial event to ensure that no contamination had occurred. These analyses were performed several times throughout the seed quality process to ensure that the seed used for commercialization was from the selected event.

**TRAIT INTROGRESSION**

After event selection, the next phase of commercial development further reduces any risk of unintended effects in the final commercial variety. Trait introgression is the process of transferring the DNA insertion (with as little as possible of the event’s genomic sequence around this insertion) into numerous different lines that have agronomic characteristics optimized for their growing region or that meet specific agronomic demands.

Most maize varieties and inbred lines are not readily amenable to genetic engineering. Therefore, inbred lines that can readily incorporate the inserted DNA are used for the transformation. The inbred lines that can be genetically engineered are typically not elite lines; that is, they are older varieties that are not commercially competitive with current elite varieties or hybrids. Therefore, the DNA insertion is transferred from the selected event into elite inbred lines or varieties through a series of backcrosses. The GM trait in any chosen event will be marketed in at least 20 to 40 commercial varieties or hybrids.
The trait introgression process can be expedited by incorporating the use of molecular markers. Molecular markers act as points of reference across the genomes of the event and the elite lines but are unique to each genome, allowing the rapid identification and selection of individual backcross progeny that have a higher proportion of elite line germplasm (Eathington et al., 2007). A typical trait introgression process includes approximately six backcross generations. The use of molecular markers can accelerate the process by facilitating the selection of progeny that contain the greatest proportion of elite line germplasm and reduce this to as few as three rounds of backcrossing. As shown in Fig. 4, during successive backcrossing using current conventional plant breeding practices (including the use of molecular markers), theoretically >99% of the event’s genome is removed from the offspring. DNA from the genetically engineered inbred that is chromosomally proximate to the DNA insertion may carry forward in the offspring (“linkage drag”), but the vast majority of the DNA in commercialized seed (and the harvested grain) is from the nontransformed, elite breeding lineage. In this regard, introgressing a transgene is analogous to backcrossing a trait from a landrace or nonadapted germplasm into an elite inbred. Therefore, these breeding processes segregate out any potential genomic mutations or epigenetic changes (variations in gene expression that are not from a genomic sequence change) that might have occurred during transformation or cell culture.

As shown in Fig. 4, during successive backcrossing using current conventional plant breeding practices (including the use of molecular markers), theoretically >99% of the event’s genome is removed from the offspring. DNA from the genetically engineered inbred that is chromosomally proximate to the DNA insertion may carry forward in the offspring (“linkage drag”), but the vast majority of the DNA in commercialized seed (and the harvested grain) is from the nontransformed, elite breeding lineage. In this regard, introgressing a transgene is analogous to backcrossing a trait from a landrace or nonadapted germplasm into an elite inbred. Therefore, these breeding processes segregate out any potential genomic mutations or epigenetic changes (variations in gene expression that are not from a genomic sequence change) that might have occurred during transformation or cell culture.

Any one cycle of trait introgression typically covers three to six generations; however, as is true for characteristics brought in through conventional breeding practices, the cycle is repeated yearly as new elite germplasm emerges from the breeding pipeline. Additionally, the DNA insertion is not integrated into the new suite of elite germplasm from the original event, but from the most recent integrated elite line with a similar genetic background. Thus, the genomic DNA around a given DNA insertion potentially becomes progressively more diluted with the DNA from the conventionally bred line at every generation of backcrossing. In the case of the DNA insertion in Table 5,
the event’s flanking DNA was no longer detectable even within 20 cM of the DNA insertion after 11 backcrosses. In this analysis, a value of 50 for the “average percentage of event DNA that is within 20 cM of the DNA insertion” would indicate that half of the DNA within 20 cM of the DNA insert is attributed to DNA from the initial transformation inbred (this analysis excludes the first centimorgan adjacent to the inserted DNA).

**PLANT SELECTION PRACTICES MINIMIZE UNSAFE UNINTENDED CHANGES**

This review of plant breeding and selection practices highlights how these methods are used to introduce beneficial characteristics into crops, whether from conventional breeding sources or from biotechnology. Importantly, these practices also minimize the advancement of unintended changes that might affect the safety of a new variety (European Commission, 2010; NAS, 2016).

When genetic engineering of plants was first being developed, it was hypothesized that this technology might induce potentially unintended changes that affect food or feed safety—for example, by activating previously dormant pathways in the plant (Kessler et al., 1992). As a result, extensive regulatory requirements for GM crops, which use a comparative safety assessment process, are now in place (König et al., 2004; Cellini et al., 2004; EFSA, 2006; Paoletti et al., 2008; CODEX, 2009; Privalle et al., 2012; Hoekenga et al., 2013; Prado et al., 2014). Since that time, numerous studies have found that GM varieties are compositionally equivalent to conventional crops (Parrott et al., 2012; Herman and Price, 2013; Hoekenga et al., 2013; Ricroch, 2013; Xu et al., 2014; Ladics et al., 2015a, 2015b; Curran et al., 2015; Venkatesh et al., 2015, 2016). The exceptions are a few cases where the desired trait confers an intended change in composition, such as improved nutrition (Chassy et al., 2008). Notably, >1300 independent global regulatory agency reviews of GM crops have concluded over the past 20+ years that, excluding GM crops with intentionally improved composition, all assessed characteristics of marketed varieties of GM crops (e.g., composition, agronomic and phenotypic) are equivalent to varieties with a history of safe use (European Commission, 2010; NAS, 2016; CLI, 2017). This is, in part, a result of the same plant selection practices being used by breeders to minimize undesirable unintended effects, whether derived from the spontaneous genetic changes that occur during conventional breeding (collectively also known as insertional effects; Schnell et al., 2015) or from the use of biotechnology to insert DNA into the plant genome.

**CONCLUDING REMARKS**

Plant breeding is a process of crop improvement that continuously addresses changing needs by introducing new genetic diversity into product portfolios. The breeding and selection processes for both conventional and GM crops provide multiple opportunities to eliminate adverse unintended effects resulting from conventional breeding and/or the transformation process. The additional rigorous molecular and phenotypic characterization of GM crops further ensures that the inserted DNA performs as intended in the crop and does not confer adverse unintended effects. Combined with an inherently low ability of genomic changes to produce harmful effects, the creation of new plant varieties has one of the safest records of all human technologies.

**Conflict of Interest**

During the conduct of this study, many of the authors (K.C. Glenn, E. Bell, M. Goley, J. Jenkinson, B. Liu, C. Martin, C. Souder, O. Sparks, W. Urquhart, and J.L. Vicini) were employees of Monsanto Company and were provided financial support in the form of authors’ salaries and research materials.

**Acknowledgments**

The authors thank Drs. Luis Burzio, Steven Halls, and Stefan Henkes for careful editorial review of this manuscript.

**Table 5. Dilution of DNA from the genome of a genetic modification event through successive backcrossing with elite varieties.**

<table>
<thead>
<tr>
<th>No. of backcrosses</th>
<th>No. of lines assessed</th>
<th>Average event DNA that is within 20 cM of the DNA insertion†</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>32</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

† DNA within 1 cM of the DNA insertion is excluded from the analysis.

**References**


