Maize Quality Assurance and Quality Control SNPs Panel

Introduction:

Maize (Zea mays L.) serves as one of the key multi-purpose crops worldwide and is utilized for food, feed, fuel, and a variety of industrial products. In the United States, most of the grain harvested is not used directly as food. The USDA estimated that ~3.5% of the grain harvested was used for cereals and other food products (USDA–National Agricultural Statistical Service, 2019). This figure reflects a threefold increase in demand and consumption of maize-based food products since the 1970s, primarily because of the popularity of both Hispanic and gluten-free foodstuffs. In Africa and Latin America, maize is a staple crop for food security and nutrition for both humans and animals. To meet increasing market demands in the United States and globally, the quality of the seed along the supply chain should be guaranteed; assessment of genetic purity of parental inbred lines and resulting hybrids is an essential measure in maize hybrid breeding and consequently, high-quality grain. Quality seed of hybrid maize is marketed when genetically pure breeder seed, basic seed, and certified seeds are produced (Poets et al. 2020).

Quality Assurance (QA) is the process used to measure and assure the quality of a product, and Quality Control (QC) is the process of ensuring that products meet consumer expectations. Seed QA and QC practices of corn were traditionally conducted using Grow Out Tests (GOT): a morphologically based approach which uses a set of morphometric descriptors, or biochemical markers analyses based on the protein profiles (isozymes) of different genotypes. In recent years, seed QA/QC has shifted to the use of molecular markers.

The molecular marker approach detects variation of the genotypes directly at the DNA level. Unlike GOT and biochemical marker methods, which have low polymorphism and high environmental influences, molecular markers are polymorphic, independent of the environment, reproducible, present at all developmental stages, probe known positions in the genome, and may be linked to traits of interest. Single Nucleotide Polymorphisms (SNPs), which are substitutions of a single nucleotide at a specific position in the genome, are currently considered to be the optimal marker type for most genotyping applications.

Large numbers of SNPs are present in all eukaryotic genomes. They are dispersed throughout the genome and are inherited as codominant Mendelian traits. The development and advancements in Next Generation DNA Sequencing (NGS) methodologies have contributed further to the usefulness of SNPs as molecular markers; SNPs lend themselves to a high level of multiplexing when they are probed using NGS. These features translate into relatively low cost and high speed of detection (Josia et al. 2021), which are two critical qualities when it comes to QA/QC applications.

The Panel:

The maize QA/QC SNPs panel was developed at the International Maize and Wheat Improvement Center (CIMMYT). For a complete description of the development of the panel, see Chen et al. (2016). The 131 SNP panel presented here is an extension of the one in the above citation.

The objective of a QA/QC SNP panel is to enable the generation of a unique SNP profile for every corn line, thus, allowing to tell any two lines apart and detect off-types within a line. The evaluation of prospective marker sets was done against 561 CIMMYT Maize Inbred Lines (CMLs). Different marker groupings were tested for their usefulness in differentiating between the different CMLs and their ability to evaluate homogeneity within the elite lines.

For the selection of the most informative marker sets, the following marker performance parameters were used:

Data completeness: Markers with more than 20% missing data were filtered out.

Minor Allele Frequency (MAF): Higher MAF improves the efficiency of distinguishing CMLs from one another. MAF was kept at 0.25 or higher.

Marker distribution: Uniform marker selection across chromosomes demonstrated a better separation of CMLs than random selection. Thus, all chromosome arms are probed with the number of SNPs per chromosome ranging from 7 to 24, and the average interval between two adjacent markers is about 17.5 million base pairs (Table 1).

Chromosome	N of SNPS	Average Interval size (BP)
1	19	15,714,358
2	21	10,869,929
3	11	22,209,162
4	11	23,522,106
5	18	12,624,620
6	14	12,151,073
7	6	32,163,306
8	16	10,902,712
9	9	17,626,338
10	6	29,133,490
Average	13.1	18,691,709

Table 1. Number of SNP markers on each chromosome and their spacing given as the average distance between two adjacent markers (BP, base pairs),

Within a finite and well-defined collection of lines, a limited number of SNPs may suffice to establish unique and diagnostic allele profiles that can distinguish each line from all others. The number of markers needed to generate such distinctive and unique DNA fingerprints will vary with the level of relatedness of the lines; the more related the lines are by common descent the higher will be the number of SNPs needed for the required resolution. The number of markers also changes during different stages of the breeding program or production stage (for example identifying inbreds vs hybrids). When the germplasm that the SNP panel is expected to provide resolution for is not defined, the total number of markers required is larger, and an even distribution of the markers throughout the genome becomes more important. The operational efficiency and cost-effectiveness of PlexSeq (see following), where there is no real cost or turnaround time penalty for increasing the number of SNPs (within the QA/QC values), allows for providing a generic global solution to all the QC/QA needs by increasing the total number of SNPs. Employing 131 SNP should enable the identification of any germplasm entry and measuring the heterogeneity within it.

The Number of Individuals in a QC Sample

The number of individuals sampled for routine QC genotyping is crucial as it affects the cost, time, and accuracy of detecting off-types within the entity. Chen et. al (2016) evaluated the influence of sample number on the probability of detecting off-types, and their findings are summarized in Fig 1. The graph captures the capacity to detect contamination in populations with defined percentages of off-types at different sample sizes. Considering the number of off-types detected in the sample from 0 to 3, detecting one off-type individual in a sample size of 192 individuals will correspond to 2% contamination at the source lot/line. If two samples were detected as off-types, the lot would have an upper limit of 2.5% contamination with 95% confidence. Therefore, if the target genetic purity of an inbred parental line is upward of 98%, the minimal size of a sample should be at least 192 individuals.

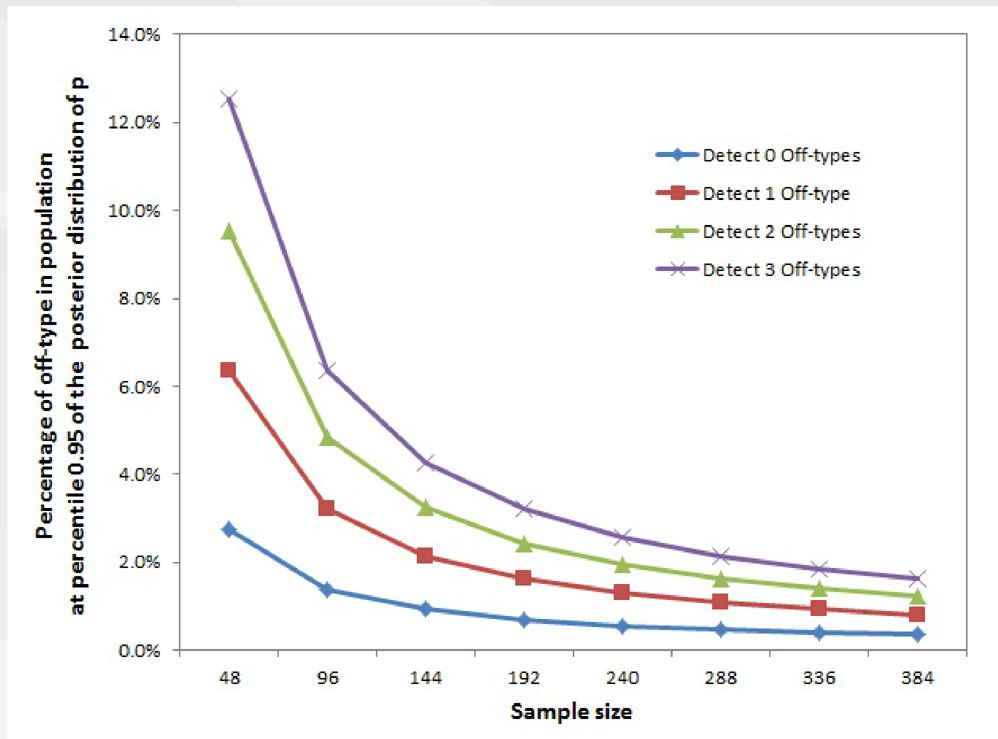

Figure 1. The capacity to detect contamination in populations with defined percentages of off-types at different sample sizes.

The Technology: *PlexSeq™:* multiplexed, SNP genotyping platform

Several attributes of the PlexSeq™ process contribute to its unique value as a genotyping platform:

- The proprietary multiplexing algorithm, *PlexForm™*: The software designs all possible primers around all requested SNPs. Artificial intelligence algorithms identify the optimal sets of primers that can be mixed in one PCR amplification reaction.
- Once the amplifications are completed, the amplicon mixture is equivalent to barcoded libraries produced from other NGS methods. The process is unique because the samples produce amplicon libraries that are equivalent in concentration and do not require any additional equalization steps. A mixture of all the libraries is subjected to one bead cleanup and is loaded onto the sequencer. The process saves time, plasticware, and expenses.
- The method requires only minute quantities of crude DNA that can be isolated from a variety of tissues, enabling a quick and inexpensive DNA isolation process to start the genotyping workflow.
- The *PlexSeq™* workflow consists of:
  - Crude DNA isolation
  - Primary PCR: highly multiplexed, low volume (3ul) PCR amplifications
  - Secondary, barcoding PCR amplifications

- Pooling: barcoded amplicons are combined into one tube, purified and quantitated
- Sequencing on an NGS sequencer
- This relatively simple workflow is amenable to automation; all steps can be carried out on liquid handlers and high-capacity thermocyclers. This enables high-throughput genotyping

Once the sequencing is complete, a proprietary allele frequency-based genotype calling analysis software, *PlexCall™*, provides an automated sequencer to data workflow. This java-based software is tuned for each assay and is fully automated based on only the sequencing output files and a sample sheet indicating the sample location on the plate.

Two other features make *PlexSeq™* a unique fit for molecular breeding and seed QA. These applications typically require the genotyping of large numbers of individuals. AgriPlex Genomics' extensive collection of barcode combinations allows for simultaneous sequencing of thousands of individuals, limited only by the sequencer's capacity. Similarly, molecular breeding may require the addition or substitution of only some of the SNP markers as the individual breeding program advances or among programs as the parent's genetic makeup or focus changes. Those changes in SNP composition are also required for QA applications as the diversity of germplasm changes. The fact that the panel is a collection of PCR primers not tethered to a surface (e.g.: chips) provides the flexibility to dynamically customize and alter the composition of the markers in the panel so it best fits the germplasm or the application. The Corn QC SNP panel is available as a service from AgriPlex Genomics and is also available as a kit to be used by in-house genotyping laboratories.

Applications:

Varietal Identification:

A variety is defined by The International Union for the Protection of New Varieties of Plants (UPOV) as: A group of plants of the same species with a common set of characteristics. These characteristics can be defined by traits resulting from given genotype(s) which distinguishes the group from other plant groupings. A variety is considered a unit suitable for being propagated unchanged.

SNPs are used for determining the unique genetic characteristics of a variety. A set panel of SNPs is used to obtain a diagnostic DNA fingerprint for the seed variety in question by genotyping a representative sample of individuals. This DNA profile, or fingerprint, is made of a set of alleles for the SNPs used in the panel that are unique to the genetic background of a given variety. The testing is done to answer the question: Do the tested individual(s) belong to a certain variety? The test is conducted by genotyping the individual(s) with the predetermined SNP panel and comparing the resulting DNA fingerprint(s) to the diagnostic DNA profile of the variety.

The requirements from the markers used for this application are that they will be:

- Polymorphic
- Have as close as possible even allele frequency (high MAF)
- Robust performance.

The number of SNPs and their genomic distribution determines the depth and resolution of the obtained DNA fingerprint. Usually, the requirement is to minimize the number of SNPs for cost efficiency. Since the cost of PlexSeq panels is less sensitive to the number of SNPs, one can use a higher number of markers for finer resolution and increased certainty.

Genetic Purity Testing

A derivative of Varietal Identification is Genetic Purity Testing. Genetic Purity is a population concept, as it assesses the genetic characteristics of a group of individuals: a line, seed lot, or seed bag, using the Varietal Identification tools.

Genetic Purity testing is set to answer the question: What proportion of all individuals of a grouping is of the desired or intended genetic make-up? The genetic make-up of each individual tested is expressed as a pattern of alleles in a given SNP panel (DNA fingerprint). Genetic Purity is documented as a percentage of the individuals that possess the expected pattern. Although, some will use the complement, which is the percentage that does not. In many cases, this proportion is tested against a threshold of relative frequency that was pre-determined based on a quality standard or by the law.

Genetic Purity testing is used by seed producers and growers as a quality control tool to identify the following: out crosses, selves (where it applies), seed mixes, and seed swaps. It is also used by regulators to check compliance.

Conclusions:

Assessment of genetic purity of parental inbred lines and resulting hybrids during breeding operations and along the supply chain is an essential measure to assure high-quality grain.

The best methodology for seed QC/QA is the use of SNPs as molecular markers; Assessing genetic purity using PlexSeq methodology in conjunction with the maize quality SNPs panel provides a generic tool fit for quality testing across the genetic width of maize.

The flexibility of the platform also enables continual revision and upgrading of the marker system, ensuring it keeps pace with the germplasm evolution.

The Maize QC SNP panel is available as a service from AgriPlex Genomics and is also available as a kit to be used by in-house genotyping laboratories.

References:

USDA–National Agricultural Statistical Service, 2019. https://downloads.usda.library.cornell.edu/usda-esmis/files/k3569432s/sj139j59z/1257b842j/cropan20.pdf

Ana Poets, Kevin Silverstein, Philip Pardey, Sarah Hearne, and James Stevenson, 2020. DNA Fingerprinting for Crop Varietal Identification: Fit-for-Purpose Protocols, their Costs and Analytical Implications. Standing Panel on Impact Assessment (SPIA), Rome

Josia C, Mashingaidze K, Amelework AB, Kondwakwenda A, Musvosvi C, Sibiya J (2021) SNP-based assessment of genetic purity and diversity in maize hybrid breeding. PLoS ONE 16(8): e0249505. https://doi.org/10.1371/journal. pone.0249505

Chen J, Zavala C, Ortega N, Petroli C, Franco J, Burgueño J, et al. (2016) The Development of Quality Control Genotyping Approaches: A Case Study Using Elite Maize Lines. PLoS ONE 11(6): e0157236. doi: 10.1371/journal.pone.0157236

The International Union for the Protection of New Varieties of Plants (UPOV) https://www.upov.int/overview/en/variety.html

Gowda, M., Worku, M., Nair, S.K., Palacios-Rojas, N., Huestis G., and Prasanna, B.M. 2017. Quality Assurance/ Quality Control (QA/QC) in Maize Breeding and Seed Production: Theory and Practice. CIMMYT: Nairobi. ISBN: 978-9966-1971-9-1