LSU 80: The Rice QA/QC PlexSeq™ SNP Panel
By AgriPlex Genomics

Introduction:

Quality Assurance (QA) is made of analyses used to measure and confirm the quality of a product. Quality Control (QC) is the process of ensuring that products meet consumer expectations. Seed QA and QC practices were traditionally conducted using Grow Out Tests (GOT): a morphology-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 (1, 2).

DNA molecular markers have evolved in the last 80 years as a means of sampling and comparing genomes. Molecular markers introduced a reliable method for QA that starts early during breeding and continues throughout seed production and as an enabler of seed trade. Over time, many DNA marker systems have been used, but Single Nucleotide Polymorphisms (SNPs) have emerged as the ultimate system; SNPs are single nucleotide changes that are heritable, codominant, and distributed with relatively high frequency throughout plant genomes. SNP markers provide the means to assess parameters such as genetic purity, varietal identification, trait confirmation, and adventitious presence without the need for grow-outs; thus, the use of SNP markers is a more rapid and cost-effective method for QA/QC compared to alternative methods (2).

Tending to the quality of rice is crucial to our ability to feed the growing human population. Rice (Oryza Sativa) is the most important food crop to human nutrition and caloric intake, providing more than 20% of the calories consumed by humans worldwide. Rice is a key agricultural commodity, being the third most produced crop worldwide (3).

Oryza sativa is a monocot of the family Poaceae (grasses) and contains two major subspecies: long-grained indica rice variety and the sticky, short-grained japonica or sinica variety. Oryza sativa japonica was first domesticated in the Yangtze River basin in China 11,500 to 6,200 years BCE, while O. sativa indica was domesticated around the Ganges River in India 6,500-2,500 years BCE. Rice has been grown in the US since the mid-19th century. Presently US rice is grown in Arkansas, Mississippi, Missouri, Louisiana, Texas, and California. In the US, this rice is purposed for direct food consumption (58%), for processed foods (16%), brewing beer (16%) and pet food (10%) (3).

We describe here a panel of 80 SNP markers that was developed at the laboratory of Professor Adam Famoso at the Louisiana State University and multiplexed in a PlexSeq™ panel by AgriPlex Genomics. We introduce PlexSeq™, a targeted amplicon sequencing methodology and possible applications for the panel as a seed QA/QC tool.

Characterizing the LSU 80 Rice QA/QC Panel

The markers in the LSU80 panel are mostly derived from the LSU500 SNP panel (4) as such they were selected based on the same marker quality criteria, where the genotyping success rate was higher than 80% and minor allele frequency higher than 0.1 (5). The panel can be divided into two marker groups.

The first group consists of 48 genome-wide markers and include markers that were previously used for US rice purity fingerprinting. The SNPs in this group are spread throughout the rice genome and were put together to distinguish any two US rice varieties from each other. The number of markers per chromosome varies from 1 – 7 SNPs (Table 1), with the average distance of 6.8 Mbp between two adjacent SNPs.

The second group of 32 SNPs includes all presently known trait associated markers relevant to the US rice. Of these, 12 SNPs are associated with various disease resistances, such as rice blast (Pita, and Pi genes) and leaf blight (CRSP2.1). Other markers are tied to fertility restoration- a trait crucial to hybrid rice production, plant and seed morphology, flavor, and cooking quality (Table 2).

| LSU80 QA/QC Rice PlexSeq SNP Panel		
Chromosome	**SNPs/Chr**	**Average Inter-marker Distance (Mbp)**
1	7	6.839613
2	6	10.001425
3	5	6.549321
4	6	5.542725
5	2	10.404459
6	4	9.813541
7	2	7.810562
8	6	9,682414
9	1	-
10	2	0.116167
11	6	5.016520
12	1	-
GW Total:	48	

Table 1. Distribution of the 48 genome wide markers along the rice genome, and the average inter-marker distance for each chromosome.

Trait Markers for LSU80 QA/QC Rice PlexSeq SNP Panel		
Markers Group	**# SNPs**	**Genes**
Eating and cooking quality	9	WaxyExons; ALK; BADH2
Disease Resistance	12	Pita2; Pi43; Pikm; Pik; Pi9; Pi33; Piz; PV; PiB; CRSp2
Fertility restoration	2	RF4
Natural Herbicide resistance	1	HIS1
Plant morphology/Growth Habit	3	SD1;
Seed/Grain morphology	5	GS3; CL; Glab1; qGL7
Trait markers: Total	32	WaxyExons; ALK; BADH2

Table 2. Traits associated markers included in the panel by phenotypic category and the genes/loci they are targeting.

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., (5) evaluated the influence of number of individuals in a sample on the probability of detecting off-types, and the summary their findings is represented in Fig 1. The graph depicts 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, 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 minimum size of a sample should be at least 192 individuals.
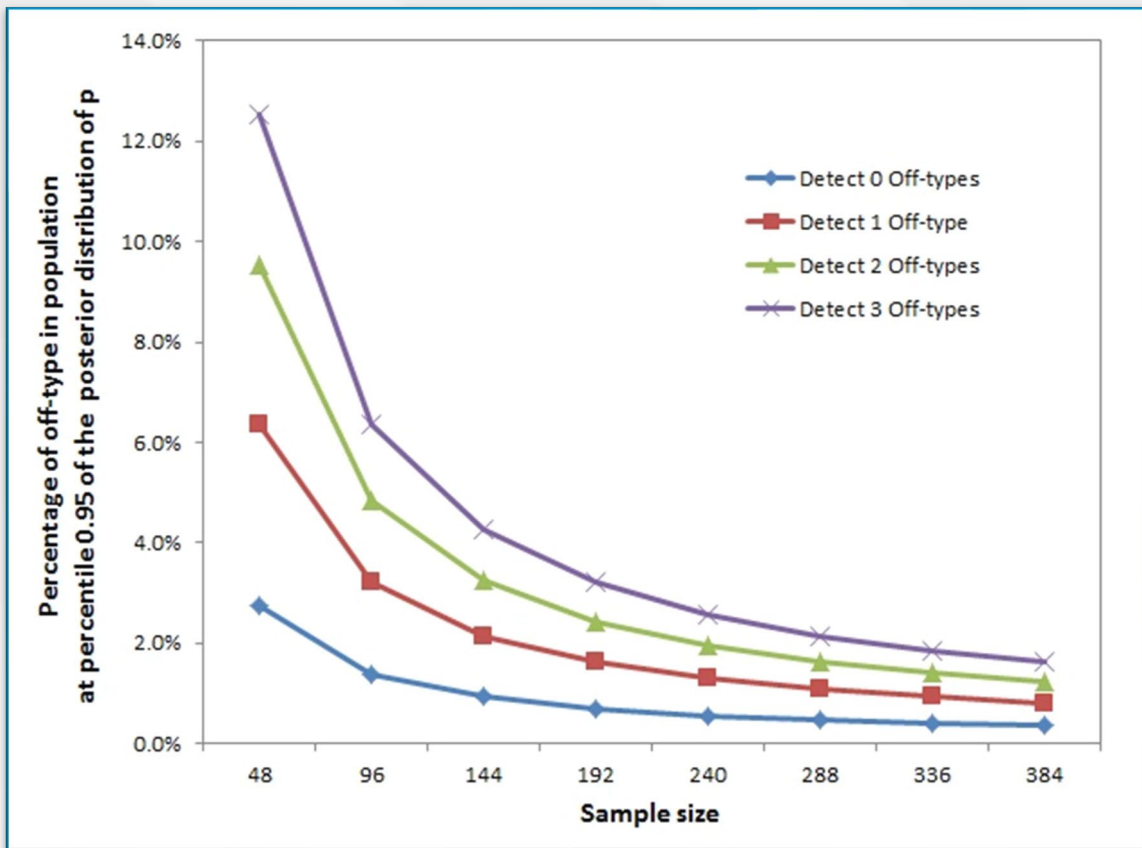
Figure 1. The capacity to detect contamination in populations with defined percentages of off-types at different sample sizes (Chen et al., 2016).

PlexSeq™: Mid-density multiplexed SNP genotyping

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, accomplished by using artificial intelligence algorithms to identify the optimal sets of primers that can be mixed into 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 DNA 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 loaded onto the sequencer. The PlexSeq procedure saves time, plasticware, and expenses.

➢ This 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 (3 ul) 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, enabling high-throughput genotyping.

Once the sequencing is complete, PlexCall, a proprietary allele frequency-based genotype calling analysis software, 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 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 number of individuals. AgriPlex Genomics' large collection of barcode combinations allows simultaneous sequencing of up to 55,000 individuals, limited only by the sequencer's capacity.

Similarly, QA applications may require the addition or substitution of SNP markers as the diversity of the germplasm changes. The fact that the panel being 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 LSU80 Panel is available as a service from AgriPlex Genomics, where the average success rate (percent genotypes called out of all possible genotypes) over a series of projects employing the panel was 93.3%. The panel and software are 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 (7). 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)
> ➢ Have 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 with 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 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 law.

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

Conclusions

The LSU80 Panel implemented at AgriPlex Genomics provides an efficient, cost-effective solution for seed QA/QC applications during breeding operations and along the supply chain. The ability to simultaneously probe an optimal number of markers across large number of samples provides a reliable snapshot of the germplasm's genetics while enabling one to address varietal identity and genetic purity of the line and confirming the presence of many traits of interest. The flexibility of the platform enables continual revision

and updating of the marker set, ensuring it keeps pace with current trait development, needs, and evolving regulatory requirements landscape.

The LSU80 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

1. 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

2. 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

3. Rice Outlook: January 2022 Nathan W. Childs, coordinator; Bonnie LeBeau, contributor, USDA- ERS, RSC-22A, Jan.14, 2022, https://www.ers.usda.gov/webdocs/outlooks/103052/rcs-22a.pdf?v=1415.1

4. LSU500: A mid-density SNP panel for the USA rice. https://www.agriplexgenomics.com/_files/ugd/ab9cc9_bcb48e0bd90f4e8aa6d0d7bd8f5bb9 d6.pdf

5. Tommaso Cerioli, Christopher O. Hernandez, Brijesh Angira, Susan R. McCouch, Kelly R. Robbins, Adam N. Famoso (2022). Development and validation of an optimized marker set for genomic selection in Southern U. S. rice breeding programs. Submitted to Plant Genome, https://doi.org/10.1002/essoar.10508975.1

6. 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

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