The 1k IRRI's Rice Custom Amplicon assay SNP Panel (1k-RiCA V.2) available through IRRI – AgriPlex collaboration

Introduction:

Molecular breeding and modern seed Quality Assurance (QA) practices require the use of molecular markers. Molecular markers have evolved in the last 80 years as a means of sampling and comparing genomes. As the technologies for interrogating genetic variation have progressed, so too has the resolution of the genomic picture the markers are able to depict.

For a breeder, the use of molecular markers in the course of breeding programs enables accurate and early selection of individuals of interest, thus, shortening the number of selection cycles needed, reducing time to market of new lines, and lowering the overall cost of breeding. To seed producers, molecular markers provide the means to assess seed quality parameters such as genetic purity, trait confirmation, and adventitious presence without the need for grow-outs; thus, it is faster and cheaper QA than traditional methods.

Over time, researchers have developed many DNA molecular marker systems, but Single Nucleotide Polymorphisms (SNPs) have emerged as the ultimate molecular marker; SNPs are single nucleotide changes that are heritable, codominant, and distributed with relatively high frequency throughout eukaryotic genomes. The desire to create high density marker chips able to interrogate large number of SNPs per DNA sample has dominated the evolution of SNP genotyping. Research has led to many high-resolution SNP arrays for rice [1, 2, 3]. Less effort has been made in developing informative, high-throughput, and cost-effective mid density genotyping solutions for applied molecular breeding programs and seed production QA. The advent of next generation sequencing technologies and genotyping by targeted sequencing provides an attractive method for mid-density SNP genotyping.

We describe here the 955 SNP markers that were developed by IRRI and multiplexed in a PlexSeq™ panel by AgriPlex Genomics. We introduce PlexSeq™, a targeted sequencing methodology, and list various possible applications for the panel as a general molecular breeding and seed QA tool.

Characterizing the 1k Rice Custom Amplicon (RiCA) v. 2 panel

IRRI designed the first version of the 1K-Rice Custom Amplicon (1k-RiCA) panel [4]. The panel was designed using Illumina's proprietary TruSeq Custom Amplicon (TSCA) 384 Index Kit workflow. The process identified SNPs that are evenly distributed across the rice genome with high quality scores and validation status. In the multiplexed AgriPlex Genomics version of the panel, an expanded set of trait markers and genetic purity diagnostic markers (Fig. 1) augmented and retained much of the genome-wide SNPs. The resulting panel combines SNPs selected from the following resources:

- 838 SNPs originated from the Cornell 6K Array Infinium Rice (or C6AIR) chip [2] and the 3,000 Rice Genomes Project [5, 6, 7]. The markers were selected based on call rate higher than 95% and high minor allele frequencies (MAF≥0.4) determined from genotypic data available on 1,172 IRRI indica rice breeding lines and indica released varieties genotyped with the C6AIR.

- 96 trait-related markers (see summary in Table 1) linked to 58 different important trait associated genes/QTLs as reported in the literature.
- 22 purity SNPs chosen for their discriminatory ability amongst elite indica material, used in determination of genetic purity, identity and hybridity.

IRRI designed this panel as a tool for interrogating genetic variation in rice varieties, offering sufficient SNP density for genomic selection, fingerprinting, and assessment of allele frequencies for 58 key traits. The panel supports breeding germplasm, classification of rice germplasm, QTL analysis for biparental populations, and assessment of background recovery in marker assisted backcrossing applications.

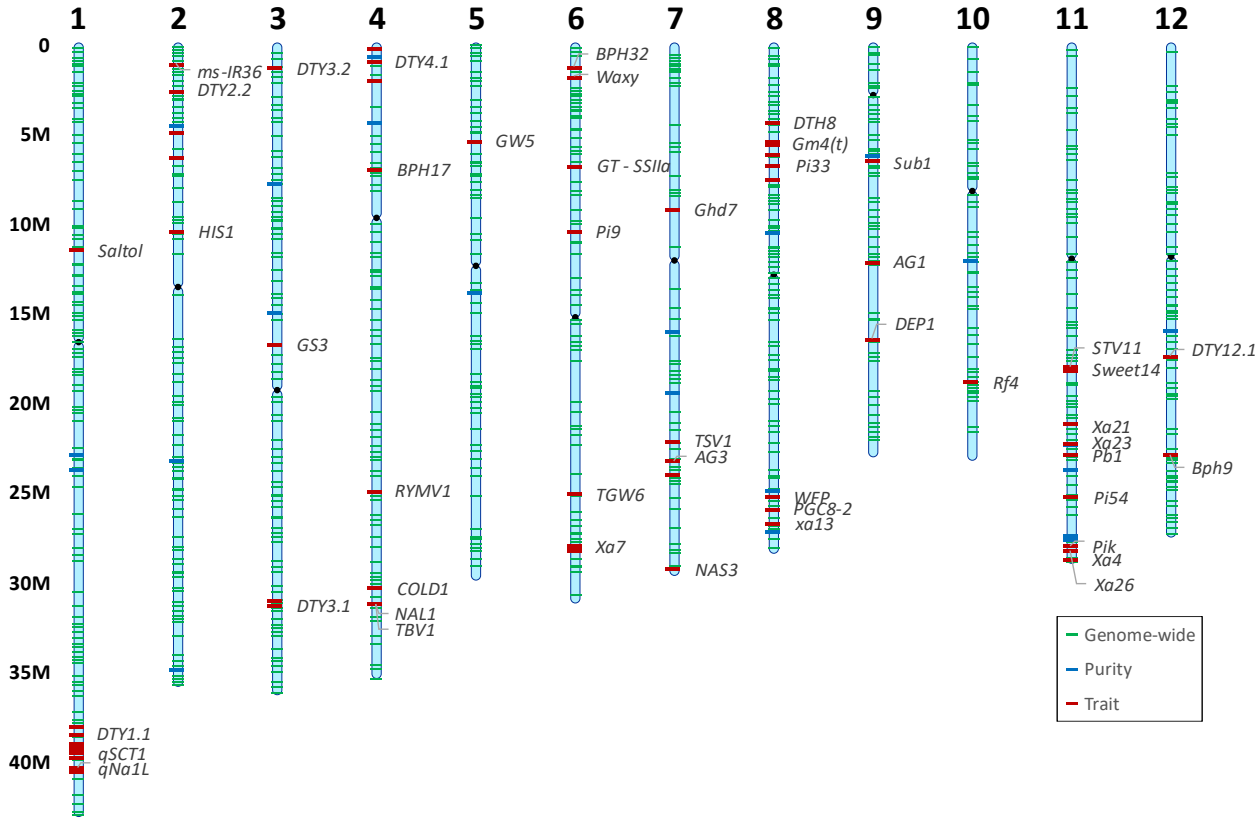

Figure 1:
Genome-wide physical position distribution of 956 SNPs from the 1k RiCA v. 2 panel across all rice chromosomes in green. SNPs designed to drive purity testing are represented in blue color, and trait-markers are in red

AgriPlex GENOMICS

PlexSeq™: The 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, and by using artificial intelligence algorithms identifies 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 loaded onto the sequencer. This 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 (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. This enables high-throughput genotyping.

Once the sequencing has completed, 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 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's large collection of barcode combinations allows simultaneous sequencing of up to 55,000 individuals, limited only by the sequencer's capacity.

Similarly, molecular breeding may require the addition or substitution of some of the SNP markers as an 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 1k RiCA V2 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

Genomic selection

The panel enables genomic selection in *indica* breeding programs. On average, any two elite *indica* parents will display between 300 – 500 polymorphic markers distributed across the genome (Figure 2), a density sufficient to enable imputation back to the full genome level. The combination of rapid turnaround time of 2 weeks and low cost enables cost-effective genotyping of a prediction population during the last generation of line fixation, saving money on the cost of field space for seed increase and allowing rapid recycling of progeny as parents.

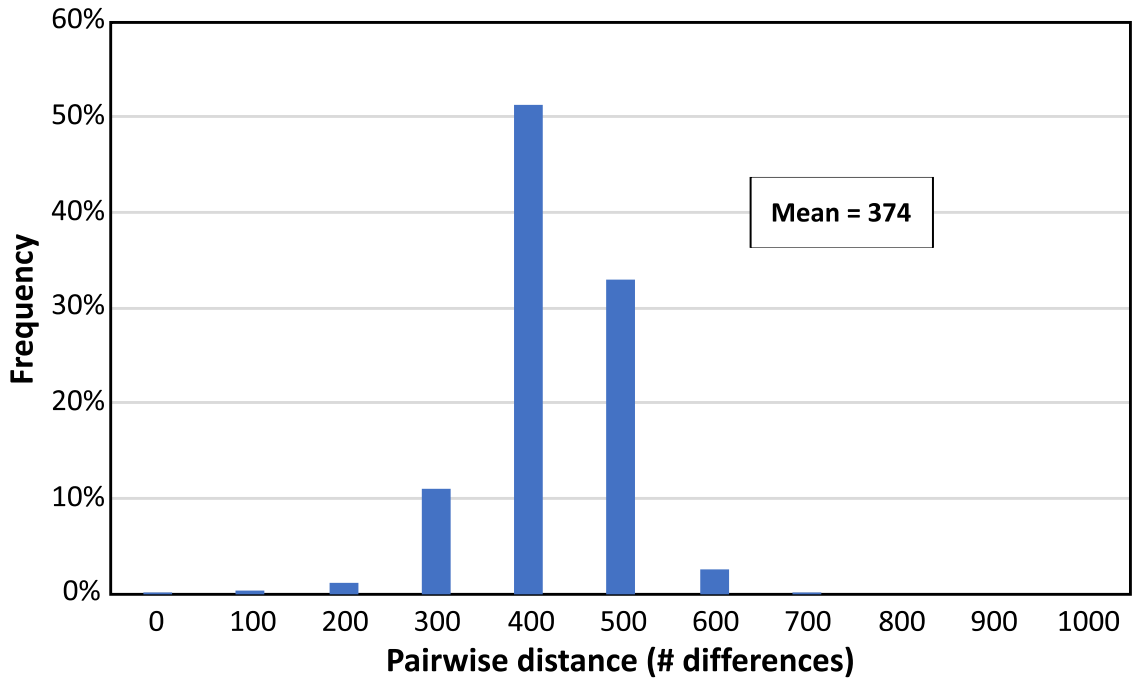

Figure 2:
Informativeness of the RiCA v. 2 panel at AgriPlex on elite indica material. The pairwise number of polymorphic markers was assessed between 87 elite indica lines from breeding programs at IRRI and elsewhere. The average number of polymorphic markers was 374, with 85% of pairs exhibiting between 300 and 500 polymorphic markers.

QTL profiling

The array at AgriPlex contains 96 trait markers targeting 58 high-value trait targets, covering a range of traits related to disease resistance, grain quality, abiotic stress tolerance, heading date, hybrid rice production and others (Table 1). These markers are designed to give accurate profiling of the targeted genes and QTLs across all *Oryza sativa* genomic diversity, including both *indica* and *japonica*. The same, or part of these QTLs and traits can be interrogated by specific, single plex assays, providing a mechanism for initial selection based on major loci followed by validation and full-genome profiling on the RiCA array.

Table 1. List of traits and genes represented on the AgriPlex RiCA panel.

Trait category	Trait	Gene targets
Disease	Blast	*Pi9, Pikh, Pi54, qPi33*
	Bacterial blight	*qXa4, xa5, qXa7, xa13, Xa21, Xa23, Sweet14*
	Brown planthopper	*Bph9, Bph17, Bph32*
	Other insects	*qGm4*
	Virus	*TSV1, TBV1, rymv1-2, rymv1-4, rymv1-5, STV11*
Grain quality	Amylose	Waxy (6 different alleles distinguished)
	Chalkiness	*PGC8.2, GW5*
	Gelatinisation temp.	*Alk-3b*
	Grain size/shape	*GS3, TGW6*
Abiotic stress	Drought	*qDTY1.1, qDTY2.2, qDTY3.1, qDTY3.2, qDTY4.1, qDTY12.1*
	Cold	*Cold1, qSCT1*
	Salinity	*qSES1.2, Saltol*
	Anaerobic germination	*qAG1, qAG3*
	Submergence	*Sub1*
Yield components	Heading date	*Ghd7, DTH8*
	Miscellaneous	*DEP1, NAL1, WFP1, HIS1*
Hybrid		*Rf4, ms-IR36, WA-CMS*

Seed purity and hybridity

The RiCA v. 2 panel includes 22 markers specifically chosen for their ability to distinguish amongst elite material. Combined with the genome-wide and trait markers, this makes the panel an excellent choice for applications related to genetic purity, some examples are, purity testing (uniformity), F_1 hybridity tests, and varietal identification identity fingerprinting. The same markers are implemented at Intertek as individual SNP assays, providing a mechanism for cross-validation of results.

Marker-assisted backcrossing

The RiCA v. 2 panel provides an excellent low-cost alternative for background recovery estimates in marker-assisted backcrossing programs. The combination of highly informative background markers together with a wide selection of peak trait markers allows accurate estimation of background recovery, ensures valuable genes from the recipient line are recovered, and in many cases provides additional confirmation that a target gene is carried by the selected progeny. Background selection can reduce the number of backcross generations required by 2 or more to achieve e.g. >95% recipient recovery.

QTL mapping

Although not the primary target application, it is possible to use the RiCA v. 2 panel for biparental mapping purposes. Density of polymorphic markers may in some cases be lower than desirable, depending on the parents involved, leading to gaps in the linkage map. However, the panel does provide a cost-effective option for genotyping much of the genome, and any remaining gaps could then be filled in with other marker systems, or by further panel customization.

Conclusions

The RiCA v. 2 panel implemented at AgriPlex provides an excellent, cost-effective alternative for applications requiring a moderate to high sample throughput at modest SNP density. While it primarily enables genomic selection, the substantially expanded suite of trait markers and the addition of purity markers expand its usefulness in a range of other applications. The panel fits neatly in with rapid line fixation protocols, thanks to its low cost-per-sample and fast turnaround time, enabling major-locus selection and genomic selection to occur before field amplification of seed, saving both time and money. The flexibility of the platform also enables continual revision and upgrading of the marker system, ensuring it keeps pace with current trait needs.

References:

1. Chen H, He H, Zou Y, Chen W, Yu R, Liu X, Yang Y, Gao YM, Xu JL, Fan LM, Li Y, Li ZK, Deng XW (2011) Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (Oryza sativa L.) Theor Appl Genet 123:869–879. https://doi.org/10.1007/s00122-011-1633-5
2. Thomson MJ, Singh N, Dwiyanti MS, Wang DR, Wright MH, Perez FA, DeClerck G, Chin JH, Malitic-Layaoen GA, Juanillas VM, Dilla-Ermita CJ, Mauleon R, Kretzschmar T, McCouch SR (2017) Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. Rice 10-40. https://doi.org/10.1186/s12284-017-0181-2
3. McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, Greenberg AJ, Naredo MEB, Mercado SMQ, Harrington SE, Shi Y, Branchini DA, Kuser-Falcão PR, Leung H, Ebana K, Yano M, Eizenga G, McClung A, Mezey J (2016) Open access resources for genome-wide association mapping in rice. Nat Commun 7: 10532. https://doi.org/10.1038/ncomms10532
4. Juan David Arbelaez, Maria Stefanie Dwiyanti, Erwin Tandayu, Krizzel Llantada, Annalhea Jarana, John Carlos Ignacio, John Damien Platten, Joshua Cobb, Jessica Elaine Rutkoski, Michael J. Thomson and Tobias Kretzschmar (2019) 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. Rice 12-55. https://doi.org/10.1186/s12284-019-0311-0
5. Alexandrov N, Tai S, Wang W, Mansueto L, Palis K, Fuentes RR, Ulat VJ, Chebotarov D, Zhang G, Li Z, Mauleon R, Hamilton RS, McNally KL (2015) SNP-seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res 43:D1023–D1027. https://doi.org/10.1093/nar/gku1039
6. Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, Sanciangco M, Palis K, Copetti D, Poliakov A, Dubchak I, Solovyev V, Wing RA, Hamilton RS, Mauleon R, McNally KL, Alexandrov N (2017) Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res 45: D1075– D1081. https://doi.org/10.1093/nar/gkw1135
7. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciangco M, Palis KC, Xu J, Sun C, Fu B, Zhang H, Gao Y, Zhao X, Shen F, Cui X, Yu H, Li Z, Chen M, Detras J, Zhou Y, Zhang X, Zhao Y, Kudrna D, Wang C, Li R, Jia B, Lu J, He X, Dong Z, Xu J, Li Y, Wang M, Shi J, Li J, Zhang D, Lee S, Hu W, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Gao Q, Niu Y, Yue Z, Naredo MEB, Talag J, Wang X, Li J, Fang X, Yin Y, Glaszmann JC, Zhang J, Li J, Hamilton RS, Wing RA, Ruan J, Zhang G, Wei C, Alexandrov N, McNally KL, Li Z, Leung H (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43–49. https://doi.org/10.1038/s41586-018-0063-9