applied biosystems

The hallmark of AgriSeq™ technology: Highly reproducible genotype calls and identification of novel genotypes

Krishna Reddy Gujjula[1], Livio Baselgia[1], Jason Wall[1], Haktan Suren[1], Prasad Siddavatam[1], Rick Conrad[1] and Jeanette Schmidt[2]

[1]Thermo Fisher Scientific, 2130 Woodward Street, Austin, TX, USA, 78744,

[2]Thermo Fisher Scientific, 3450 Central Expressway, Santa Clara, CA, USA, 95051

Introduction

Amplicon based targeted Genotyping-By-Sequencing (GBS) has enabled scientists to screen for known and novel variants in a high throughput and cost effective way. AgriSeqTM GBS technology provides greater flexibility to study multiple variant types (SNPs, MNPs, INDELs and structural variants) in a single genotyping panel.
We evaluated the performance of AgriSeq™ technology for genotype call rate, concordance between replicates, and ability to identify novel variants within the targeted regions. We assessed six GBS panels ranging from 377 to 5736 markers that represent both animal and plant kingdoms.

AgriSeq™ workflow

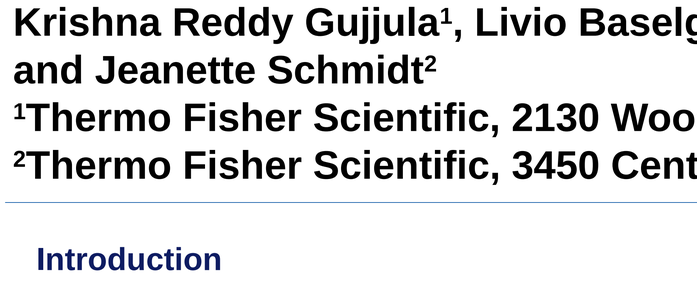

Figure 1. Workflow for performing sequencing and analysis

Dataset

Table 1. Data used in this study

Organism	# of markers	# of samples (Multiplex)	Replicated (Yes/No)
Tomato	5736	334 (~96x4)	No
Canine	387	96 (~96x2)	Yes
Sugarbeet	1055	384 (~384x4)	Yes
Cucumber	3044	96 (~96x2)	Yes
Cacao	1060	174 (~96x2)	No
Salmon	3152	384 (~96x4)	No

Materials and Methods

Six genotyping panels were designed using a reference-based GBS automated pipeline, which selects primers based on optimized parameters such as amplicon length, melting temperature and GC content, for multiplexing 100s to 1000s of oligonucleotides in a single PCR reaction. Primers were also designed to avoid overlapping known SNPs and prevent the formation nonspecific PCR products, characteristics which were assessed *in-silico*.
The panels were tested using the AgriSeq™ GBS workflow using either the 96 or 384-well AgriSeq™ HTS Library prep protocol with 10 ng of genomic DNA input. Barcoded amplicon libraries were pooled 1:1 and loaded onto an Ion Chef™ System for template prep and chip loading onto an Ion 540™ chip, and then sequenced on an Ion S5™ XL System.
Data were analyzed using the Torrent Variant Caller plugin available as part of the Torrent Suite software package, to determine the genotype calls.

Results

- All samples are from **FIELD** (not control). Some of the samples had lower than recommended input DNA.
- Sample call rate: 90-95%
- Marker call rate: 90-95%
- Sample marker concordance: ~99%

- Sample call rate: The percentage of markers for a particular sample that generate a genotype call (Fig. 2).

- Marker call rate: The percentage of samples for a particular marker that generate a genotype call (Fig. 3).

- Sample concordance: The percentage of markers that give the same genotyping call across sample replicates (Fig. 4).

Sample call rate; 174 to 1512 samples

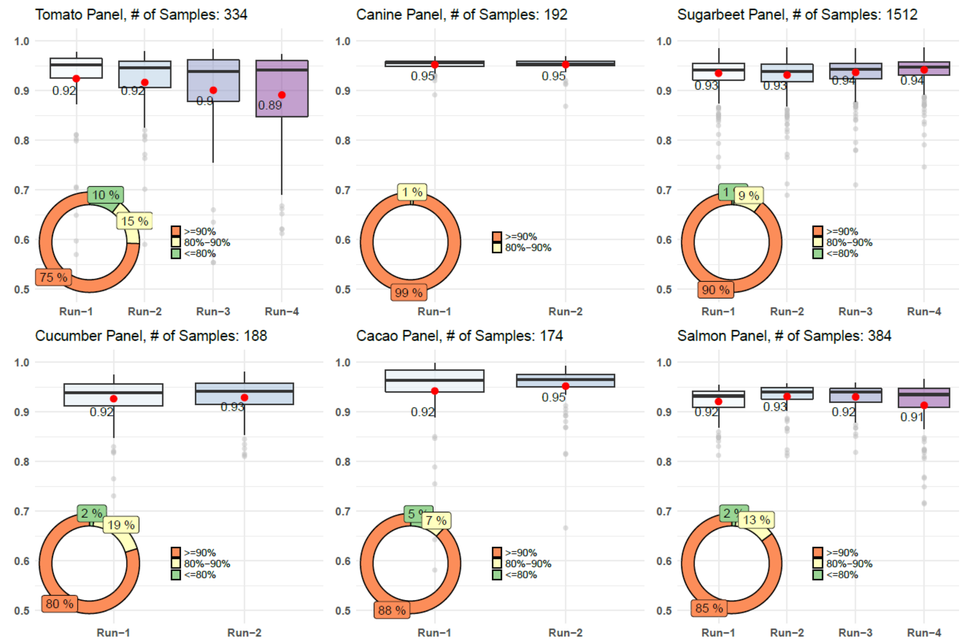

Figure 2. Shows the distribution of sample call rate for each run in each individual panel (6 panels). Ring plot for each panel classifies samples based on the call rates (>90%, 80-90%, <80%).

Marker call rate; 387 to 5736 markers

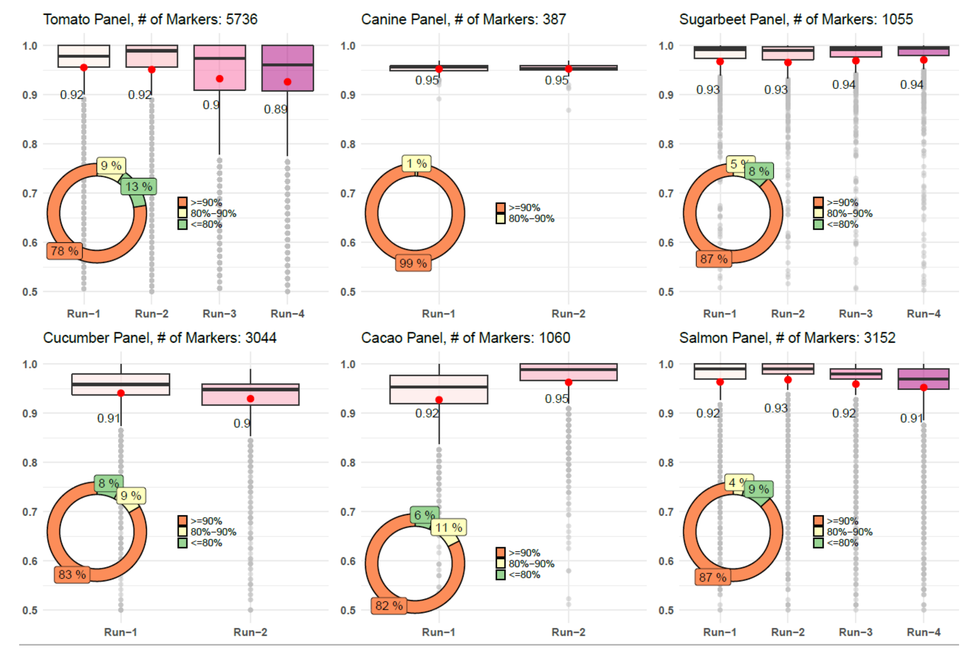

Figure 3. Shows the distribution of marker call rate for each run in each individual panel (6 panels). Ring plot for each panel classifies markers based on the call rates (>90%, 80-90%, <80%).

Are the Marker calls Concordant?

Figure 4. Concordance of marker calls between replicated samples. Each point represents a replicated sample.

- High concordance shows the high reproducibility of results in spite of several sources of variations/noise.

Does AgriSeq™ call Novel genotypes? Yes!

Table 2. Shows the number of novel markers called by AgriSeq™

Organism	# of designed markers	# of amplicons	Avg. # of Novel Markers/Sample
Tomato	5736	5212	1195
Canine	387	371	140
Sugarbeet	1055	1041	1251
Cucumber	3044	3013	5011
Cacao	1060	1054	540
Salmon	3152	3148	1235

Note: Novel genotypes are called only on the designed amplicons.

Are these Novel genotypes biologically relevant?

Table 3. Shows the support for novel markers in publicly available databases (NCBI and ENSEMBL)

Organism	dbSNP info available ?	Proportion of Novel genotypes in dbSNP (P1)	Proportion of randomly generated genotypes in dbSNP (P2)	(P1 > P2) Test on proportions p.value
Tomato	Yes	0.75	0.023	<<0.01
Canine	Yes	0.73	0.018	<<0.01
Sugarbeet	No	NA	NA	NA
Cucumber	No	NA	NA	NA
Cacao	No	NA	NA	NA
Salmon	Yes	0.30	0.015	<<0.01

- A much higher proportion of AgriSeq™ novel genotypes were supported by SNP databases than the SNPs which were generated randomly.
- The SNP positions for benchmarking were randomly generated from the targeted regions (amplicons).
- NA – Known SNP data is not available in the SNP databases.

Are the Novel calls Concordant?

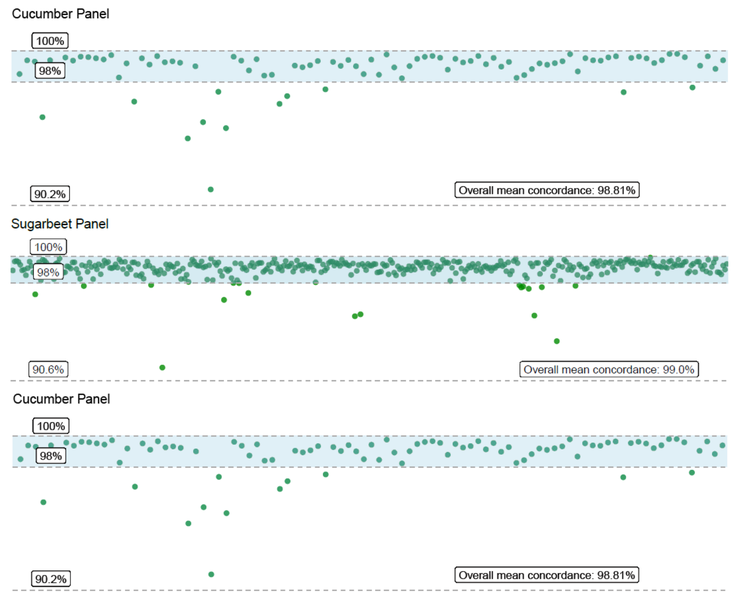

Figure 5. Concordance of novel genotype calls between replicated samples. Each point represents a replicated sample. High concordance for novel calls shows robustness and reproducibility of AgriSeq™

Conclusion

The ability to design panels ranging from fifty to several thousand genotypes for both animals and plants shows the versatility of the AgriSeq™ technology. The high concordant calls for hotspots and novel genotypes indicate that the AgriSeq™ technology results are highly reproducible while maintaining high call rates. The high reproducibility of results shows the robustness of AgriSeq™ technology in spite of several sources of variations such as sample heterogeneity, library prep, sequencing runs and human error while simultaneously genotyping known markers and discover biologically relevant novel markers.

TRADEMARKS/LICENSING

ThermoFisher SCIENTIFIC