applied biosystems

Development of targeted GBS panels for breeding and parentage applications in cattle and swine

Angela Burrell[1], Prasad Siddavatam[1], Michelle Swimley[1], Chris Willis[1], Maarten de Groot[2], Ryan Ferretti[3], and Rick Conrad[1], [1]Thermo Fisher Scientific, 2130 Woodward Street, Austin, TX, USA, 78744. [2]VHL Genetics AgroBusiness Park 100, 6708 PW Wageningen, Netherlands. [3]Neogen GeneSeek Operations, 4131 N. 48th St., Lincoln, NE 68504

ABSTRACT

Parentage testing and genomics-assisted breeding are critical aspects of successful herd management. Due to its highly accurate and reproducible results, targeted GBS is becoming an increasingly favored technology for SNP genotyping. With the utilization of next-generation sequencing, labs can test hundreds of samples across thousands of SNPs simultaneously in a simple high throughput workflow starting from either extracted nucleic acid or crude lysis samples.

We developed targeted sequencing panels for both cattle parentage, based on 200 SNP markers selected by the International Society of Animal Genetics (ISAG), and swine breeding using a 1500 SNP imputation panel. Utilizing the AgriSeq™ HTS Library Kit, a high-throughput targeted amplification and re-sequencing workflow, each panel's performance was tested on >96 diverse cattle and swine DNA samples. Libraries were sequenced on the Ion S5™ using an Ion 540™ chip with genotyping calling generated using the Torrent Variant Caller (TVC) plugin

The mean genotype call rate of markers across the samples was >98% for the cattle panel and >96% for the swine panel. Concordance across replicate library preparations and independent sequencing runs was >99.9% for both panels. Panel results were compared with results from a DNA array and the genotype call concordance was >99% with the AgriSeq workflows. The cattle panel was also used on field samples by a Netherland service lab to successfully determine the parentage relationships of 45 calves with 48 potential mother cows.

The data demonstrates the utility of the AgriSeq targeted GBS approach for cattle and swine SNP genotyping applications.

INTRODUCTION

Production agricultural applications require consistent genotyping performance and high marker call rates to ensure accurate selection. Unlike non-targeted GBS approaches (e.g. RADSeq) that are highly susceptible to allele drop outs and missing data, AgriSeq targeted GBS are designed to deliver reproducibly high marker call rates across diverse sample sets.

The AgriSeq workflow is a high-throughput, targeted GBS workflow designed to amplify and sequence thousands of genetic markers in a single multiplexed reaction. The workflow can be automated on most standard liquid handling platforms for decreased hands-on time and increased throughput. The AgriSeq workflow was tested with the Bovine ISAG SNP Parentage Panel (2013), a panel of primers targeting the 100 core and 100 additional SNPs as standardized by the International Society of Animal Genetics (ISAG) and a 1500 marker pig genotyping panel.

AgriSeq data is highly reproducible and concordant to orthogonal genotyping technologies, such as DNA microarrays. Unlike non-sequencing based approaches, AgriSeq can also help discover additional novel variants in the amplicons of the targeted SNPs. Additional variants and microhaplotype information can provide new markers for linkage analysis or enhance discrimination in parentage and traceability applications.

MATERIALS AND METHODS

The Bovine ISAG SNP Parentage Panel (2013) performance was tested on 115 diverse bovine DNA samples. Samples included a panel of 96 samples obtained from the USDA (MARC Beef Cattle Diversity Panel v2.9), as well as samples contained within the 2015 ISAG/ICAR 3rd SNP Typing Bovine Comparison Test panel. Samples originated from 20 diverse bovine breeds. The Porcine SNP panel was tested with 96 diverse pig DNA samples from North America obtained from a collaborator. The libraries were prepared using the AgriSeq HTS Library Kit (Figure 1).

Figure 1. AgriSeq Library Prep workflow

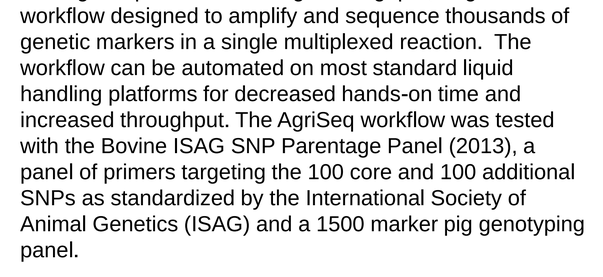


Figure 1. In the first step, 10ng/rxn of porcine or bovine DNA was amplified using the appropriate GBS panel. Each sample was treated with a Pre-ligation Enzyme to remove residual primer dimers allowing for more efficient sequencing. Samples were ligated with unique barcoded adapters allowing them to be pooled for subsequent clean-up and sequencing while retaining traceability to the original sample during analysis. N=16 samples were pooled after barcode ligation allowing for simplified sample handling and decreased cost per sample. Libraries were cleaned-up by a two-round AMPure purification. A final bead-based normalization step helps ensure each library is at a consistent final concentration suitable for direct input into template prep on the Ion Chef™ instrument. All libraries were pooled 1:1 for sequencing in a single reaction.

Libraries were sequenced on a single run on the Ion S5™ sequencing system using an Ion 540™ chip. Utilizing this system, up to 768 samples can be barcoded and run on a single sequencing run allowing for up to approximately 1500 samples to be tested a day (2 runs/day). Data was analyzed using the Torrent Variant Caller (TVC) plugin as part of the Torrent Suite™ software package to determine the genotype call for each marker and sample (Figure 2).

Figure 2. AgriSeq Sequencing Workflow

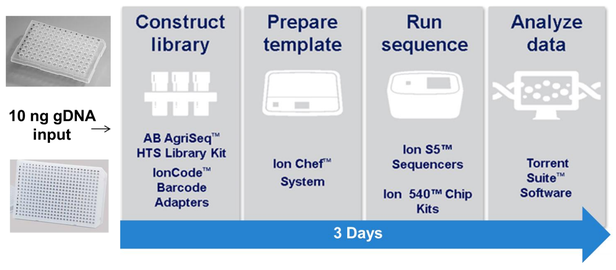


Figure 2. The complete workflow (from DNA to results) can be completed in as little as three days. Libraries are constructed the AgriSeq HTS Library kit in either 96-well or 384-well format. Template prep is performed on the Ion Chef™ System and samples are sequenced on the Ion S5™ XL System. Data is automatically analyzed using the Torrent Variant Caller plugin to generate genotype calls for all markers tested.

Call rate, the percent of markers generating a genotype call for a specific sample, was calculated for all samples tested. Mean call rates were determined across sample technical replicates as well as a grand mean across all samples to evaluate panel performance (Figures 3 and 4). Genotype concordance, the percent of genotype calls across two or more replicate samples that are identical, was also determined as well as concordance to an orthogonal method, microarrays. Potential additional novel SNPs were identified that fell within the amplicon

RESULTS

Figure 3. Bovine Panel Mean Sample Call Rate

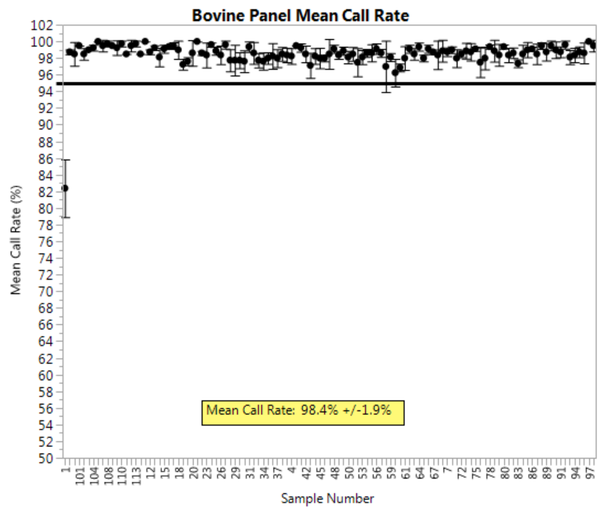


Figure 3. For the bovine panel, an excellent mean call rate (98.4% +/-1.9%) was observed during panel testing. 114 out of 115 samples had mean call rates above 95%.

Figure 4. Porcine Panel Mean Sample Call Rate

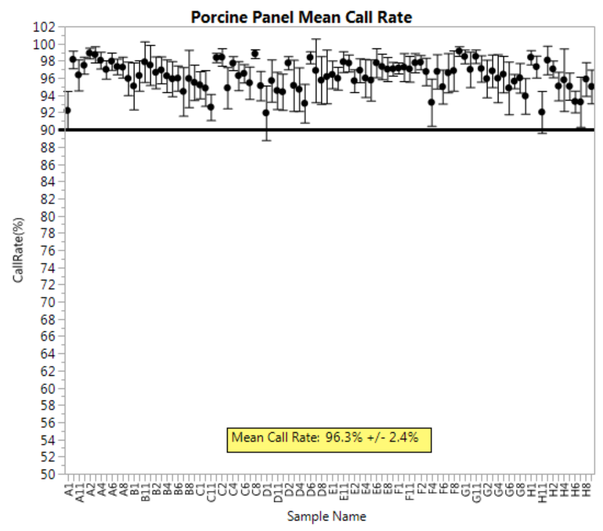


Figure 4. For the porcine panel , an excellent mean call rate (96.3% +/-2.4%) was observed during panel testing. All 96 samples had mean call rates above 90%.

Figure 5. Mean Read Uniformity

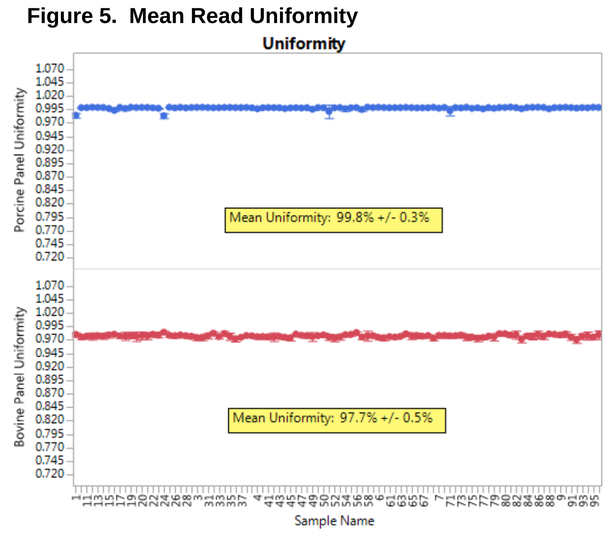


Figure 5. Read uniformity is the percentage of target bases covered by at least 0.2X of the average base read depth. It is a measure of how evenly you are covering target amplicons with reads. Low uniformity (<90%) can lead to marker drop-off and poor call rates. The mean read uniformity for both panels was excellent. The Bovine ISAG SNP Parentage Panel (2013) had a mean uniformity of 97.7% and the Porcine panel mean uniformity was 99.8%.

Figure 6. Call Rate Across Breed Comparison

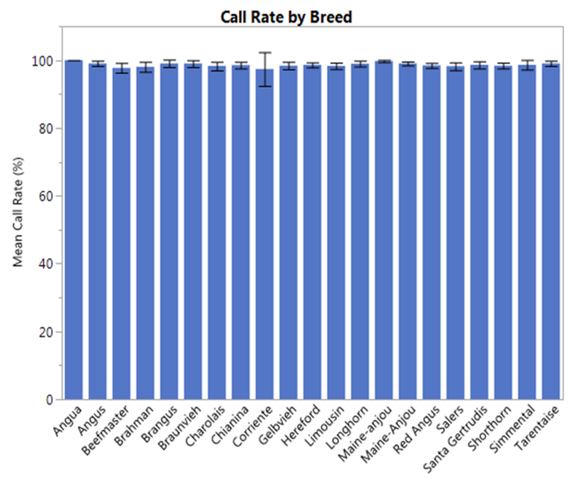


Figure 6. An ANOVA was performed to determine whether there was a statistical difference in call rate between the 20 different bovine breeds tested. There was not a statistically significant difference in call rate between the different breeds.

Table 2. High Replicate Genotype Concordance

Panel	Metric	Mean Genotyping Call Concordance	Stdev
Bovine ISAG	Intra-run Concordance	99.96%	0.13
	Inter-run Concordance	100.00%	0.00
PICv2 1500	Intra-run Concordance	99.94%	0.05
	Inter-run Concordance	99.90%	0.00

Table 2. To determine genotype concordance we ran a single sample in 96 replicates through the AgriSeq workflow for both the bovine and porcine panels. The experiment was repeated on two separate days. Library sets were sequenced on different sequencing runs. Genotype concordance was calculated between the replicates on the same sequencing run and between replicates on separate sequencing runs to determine intra-run and inter-run genotype concordance. Both inter-run and intra-run concordance was >99.9% for all panels.

Table 3. 2015 ISAG/ICAR Parentage Results

	Thermo Fisher Scientific Results	2015 ISAG/ICAR Reported Results	Results
Bovine 9 and 21	Parentage Verified	Parentage Verified	✓
Bovine 17 and 21	Parentage Excluded	Parentage Excluded	✓

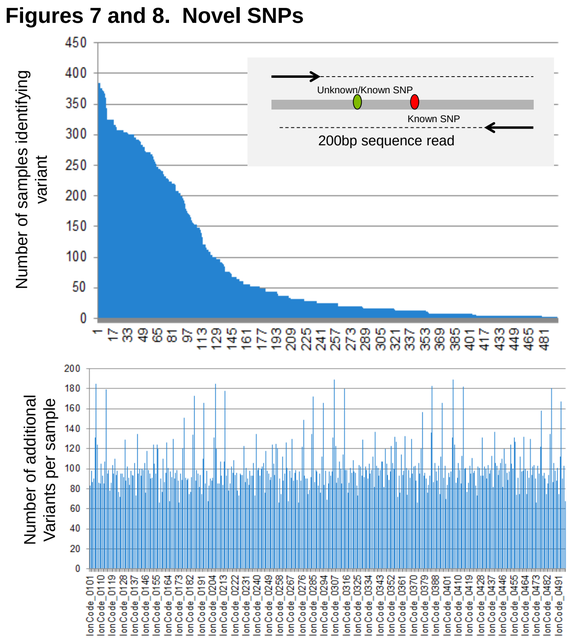
Table 3. Two pairs of samples in the 2015 ISAG/ICAR 3rd SNP Typing Bovine Comparison Test were tested for parentage validation and compared to the reported results. Using the AgriSeq workflow and ISAG criteria for parent verification (July 2012), both pairs of samples were tested. Thermo Fisher Scientific results were completely concordant with the 2015 ISAG/ICAR reported results.

Table 4. Concordance with Microarray Data

Samples run on both technologies	44
Total number of markers	8800
Number of concordant calls	8751
Concordance (%)	99.4%

Table 4. Bovine samples were hybridized to six microarrays in order to obtain consensus genotype calls for the microarray data. Concordance was calculated as the number of times the genotype call matched between samples run on the two different technologies divided by the total number of markers (200).

Figures 7 and 8. Novel SNPs



Figures 7 and 8. Since AgriSeq is based off next-generation sequencing technology, it has the advantage of being able to identify additional SNPs that fall within the targeted amplicon regions. Additional SNPs can be utilized for linkage analysis and the generation of microhaplotypes for improved marker specificity or enhanced discrimination in parentage/traceability analysis. Figure 7 shows 495 additional variants identified in the USDA gDNA samples at varying frequencies (between 4-384 samples). An average of 101 additional variants were identified across samples (Figure 8).

In a service lab field test 96 animals were tested with the AgriSeq workflow and Bovine ISAG panel (45 calves and 48 potential mother cows). A mean expected heterozygosity of 0.45 and mean polymorphic information content (PIC) of 0.35 was found. The lowest observed heterozygosity (HO) was observed for marker 90561_946 (0.10), while the highest HO was observed for EF03408 (0.65). Based on the results it was concluded that of the 45 calves, 34 could be linked to a single mother. The remaining 11 calves could not be linked to a cow. This could suggest that either the mother was not present on the farm anymore or the calf was (illegally) imported to the stable. The case study proves that the Bovine ISAG SNP Parentage Panel (2013) utilizing the AgriSeq workflow can be used in parentage verification cases.

CONCLUSIONS

The AgriSeq library prep workflow along with custom GBS panels provide a streamlined, cost-effective method for bovine parentage verification and animal genotyping. Up to 4X 384-well plates can be processed in a single day and full sequencing results can be obtained in as little as three days. The flexibility of AgriSeq allows hundreds of samples to be pooled together into a single sequencing run targeting hundreds to thousands of markers.

Our method yields calls for the vast majority of markers (mean 98.4% for the bovine panel and 96.3% for the porcine panel) with no bias between bovine breeds. Run-to-run concordance is >99.9% and calls were highly concordant with orthogonal microarray data (99.4%). While we demonstrated the utility of Ion Torrent sequencing technology for assessing parentage in cattle and genotyping in pigs, our approach can be applied to other agricultural genotyping problems as well.

In conclusion, the AgriSeq library prep kit and custom GBS panels combine into a robust and efficient workflow for animal genotyping and parentage applications.

REFERENCES

1. Cariou M, Duret L, Charlat S. How and how much does RAD-seq bias genetic diversity estimates? BMC Evolutionary Biology. 2016;16:240. doi:10.1186/s12862-016-0791-0.

2. Kidd, Kenneth K. et al. Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics . Forensic Science International: Genetics , Volume 12 , 215 – 224.

ACKNOWLEDGEMENTS

We would like to acknowledge Michael Heaton at the USDA for supplying the MARC Beef Cattle Diversity Panel v2.9 used in this study and Louis Cancela for supplying the 2015 ISAG/ICAR DNA and reported results.

TRADEMARKS/LICENSING

ThermoFisher SCIENTIFIC