Proc. Assoc. Advmt. Anim. Breed. Genet. 22:517-520

DEVELOPMENT OF A CUSTOM ION AGRISEQ GENOTYPING-BY-SEQUENCING PANEL BASED ON THE ISAG BOVINE CORE PARENTAGE MARKERS

Prasad Siddavatam[1], Angela Burrell[1], Ryan Ferretti[2], Adam Allred[1] and Kristian Ridley[3]

[1]Thermo Fisher Scientific, 2130 Woodward Street, Austin, TX, 78744 USA,
[2]Neogen GeneSeek Operations, 4131 N. 48th St., Lincoln, NE 68504 USA
[3]ThermoFisher Scientific, 5 Caribbean Drive, Scoresby, Victoria, Australia

SUMMARY

Historically, microsatellites have been the most popular genetic feature for distinguishing cattle breeds for the purpose of determining parentage. More recently, SNP genotyping has emerged as a desirable alternative to microsatellite typing. SNPs offer several advantages over microsatellites. Perhaps the most important advantage is that there is less ambiguity in distinguishing SNP alleles in order to confidently provide a genotype call.

AgriSeq™ is a sequencing technology that can be used for targeted amplification and re-sequencing of thousands of SNP targets in a single reaction. The Ion 540™ chip allows hundreds of samples to be genotyped at thousands of loci simultaneously. Ligating a unique barcode to each sample allows samples to be sequenced together in a single run on the Ion S5™ sequencing system.

We developed a targeted sequencing panel based on 200 bovine SNP markers selected by the International Society of Animal Genetics (ISAG) for the purpose of determining parentage. We tested this panel on 96 bovine samples obtained from the USDA representing 19 different breeds. Each sample was tested in duplicate such that 192 libraries were pooled onto a single Ion 540 chip for sequencing. Variant calling was performed using the Torrent Variant Caller (TVC) plugin as part of the Torrent Suite™ software package. Mean call rate for this dataset was 98.5%, indicating that the vast majority of SNPs yielded data of sufficient quality to make a genotype call.

INTRODUCTION

SNPs are well-suited for use as genetic markers for several reasons. Some of the advantages of using SNPs relative to other types of genetic markers are that SNPs occur abundantly in the genome, are generally stable through evolution and have a low mutation rate (1).

SNP genotyping has various applications in agriculture including genetic diagnostics, germplasm identification and genomic selection for breeding purposes (1). Next-generation sequencing allows for rapid and accurate SNP genotyping. This technology, coupled with the specificity of targeted amplification using AgriSeq, enables many samples to be genotyped simultaneously without compromising sensitivity.

Here we apply SNP genotyping for assessment of bovine parentage. Using our targeted sequencing primer design pipeline, we designed primers for the amplification and subsequent sequencing of 200 SNPs related to bovine parentage. The resulting panel was tested on 96 bovine samples representing 19 different breeds of cattle in order to assess call rate and concordance with array-based genotyping methods.

MATERIALS AND METHODS

The bulk of the primers were designed using an automated process that optimizes a number of oligonucleotide properties (GC content, melting temperature, etc.) and amplicon properties (size, centering a SNP within its amplicon, etc.). Furthermore, primers were designed to avoid

overlapping nearby SNPs and are subjected to an in-silico quality assessment to ensure specificity within the genome and prevent the formation of undesired PCR products.

Library prep was performed in duplicate for each of 96 samples obtained from the USDA MARC Beef Cattle Diversity Panel v2.9. All 192 libraries were pooled onto a single Ion 540 chip for template prep and sequencing on the Ion Chef™ and Ion S5 XL.

RESULTS AND DISCUSSION

Sequencing yielded over 71 million reads. Over 66 million reads were reliably assigned to one of 192 barcodes (an average of over 347,000 reads per barcode) with >99 percent of bases aligning to the bovine genome (Table 1.)

Table 1. Sequencing summary

Number of samples	96
Number of markers	200 (ISAG)
Sequencing time	2.5 hours
Analysis time	<17 hours
Output file size	10 MB (VCF)
Mean read length	140 bp
Total reads (high quality)	>66 million
Percentage of reads mapped	99.50%

Marker coverage was highly consistent. Mean coverage was 368.6 with 97% of markers falling within one standard deviation of the mean (Figure 1.).

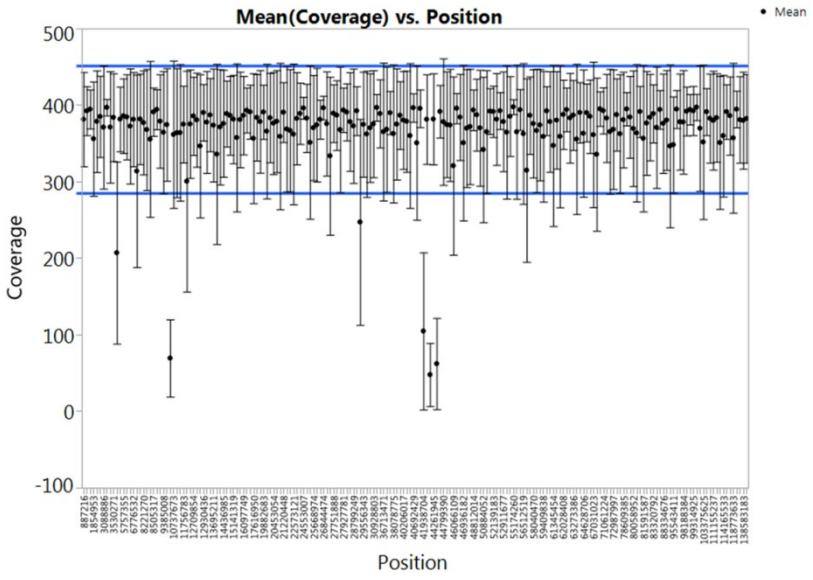

Figure 1. Marker coverage

Average call rate for these samples was 98.5%. Call rates are color-coded by breed and differences between breeds were not found to be statistically significant. Call rate was calculated for each sample as the number of markers for which data quality was high enough to make a

genotype call (homozygous reference/homozygous variant/heterozygous), divided by the total number of markers in the panel (200). Seven replicates that had less than 100x coverage were excluded from analysis (Figure 2.)

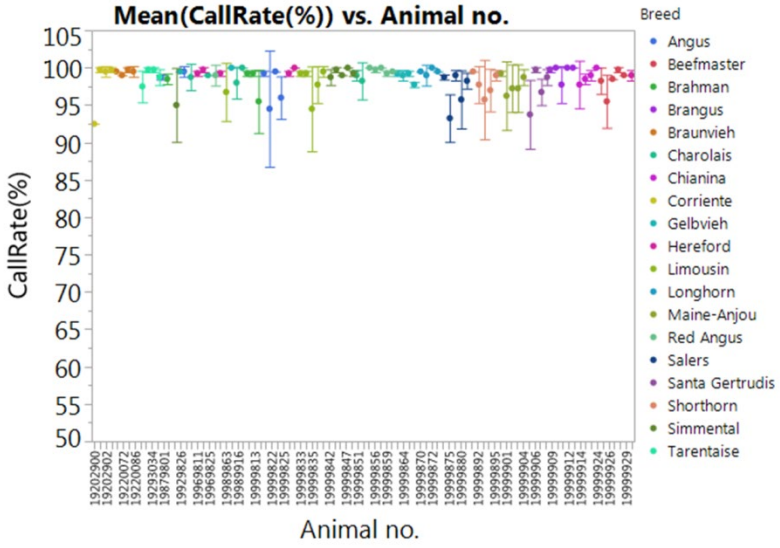

Figure 2. Sample call rates

Average marker call rate was 98.5%. 192 of the 200 markers had call rates >95% and 49 markers had 100% call rates. Only five markers had call rates <90% (Figure 3.).

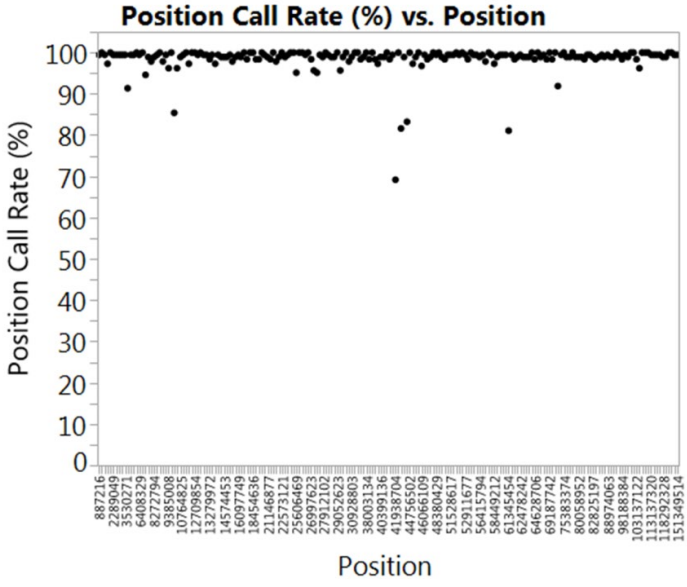

Figure 3. Marker call rates

Further investigation of the five markers that had lower call rates (<90%) revealed that performance for those markers differed greatly depending on the breed (Figure 4.).

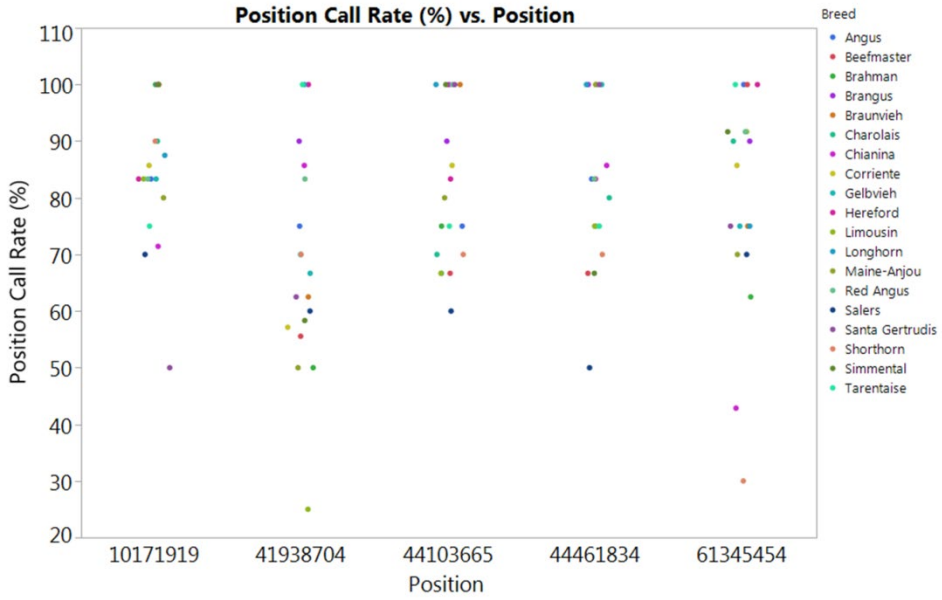

Figure 4. Markers with lower call rates show breed-specific differences in performance

Samples were hybridized to six Illumina arrays in order to obtain consensus genotype calls for the array 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 calls. Seven replicates that had less than 100x coverage were excluded from analysis (Table 2.).

Table 2. Concordance with array data

Samples included in analysis (>100x coverage)	89
Total number of calls	36399
Number of concordant calls	35433
Concordance (%)	97.3

CONCLUSIONS

We developed a high-performing, high-throughput method for genotyping hundreds of bovine samples in a single sequencing run at hundreds of SNPs. Our method yields calls for the vast majority of markers (98.5% on average). These calls were highly concordant with array data (97.3%). While we demonstrated the utility of Ion Torrent sequencing technology for genotyping parentage markers in cattle, our approach can also be applied to other SNP genotyping problems.

REFERENCES

Patel DA, Zander M, Dalton-Morgan J, Batley J. (2015) Methods Mol Biol. 1245:1-11.