A customizable high-throughput genotyping technology that permits fast, easy, and inexpensive alteration of marker content

Claudio Carrasco, Jason Wall, Angela Burrell, Krishna Reddy Gujjula, Hakten Suren, Prasad Siddavatam, and Christopher C. Adams

Thermo Fisher Scientific, 2130 Woodward Street, Austin TX 78744

INTRODUCTION

High-throughput genotyping solutions for parentage and breeding applications require the ability to simultaneously interrogate hundreds to thousands of genetic loci both easily and economically. One disadvantage of many high-throughput genotyping technologies is the lengthy lead times and considerable cost associated with changing the genomic marker content (targeted loci) in an assay. The Applied Biosystems™ AgriSeq™ targeted genotyping-by-sequencing (tGBS) solution for plant and animal genotyping does not suffer from this problem because the technology relies on a pool of PCR oligonucleotides that can be quickly, easily, and inexpensively changed to accommodate an always improving knowledge of genomic function. If and when the need arises to alter the content of a marker panel, all that is required is the design and synthesis of additional PCR primers, which are then simply spiked into an existing assay pool. In addition, AgriSeq genotyping panels can be ordered in plate format in which primer pairs for marker-containing amplicons are individually aliquoted, enabling the user to remove unneeded amplicons or re-formulate primer pools (panels) in any combination desired. Furthermore, individual panels targeting specific species can mixed together, creating a multi-species panel while still enabling species-specific genotyping. Here we show that a mixture of three mid-density panels for multiple species not only allowed for accurate species-specific genotyping, but also enabled the accurate assignment of species to unknown gDNA samples being tested. This unparalleled flexibility in a highly multiplexed genotyping platform provides users unlimited avenues for customizing their genotyping workflows.

Introduction

Increasing pressures in plant and animal production drive the need for more efficient marker–assisted breeding approaches and methods for fast and accurate parentage identification. Targeted Genotyping- by-Sequencing (tGBS) uses massively parallel PCR and next-generation sequencing for association studies at unrivaled marker and sample densities. The AgriSeq™ approach highlighted here provides sequence information for the target markers free of prediction (see overview of AgriSeq targetd GBS workflow below in Figure 1). Some of the advantages of this approach over alternative genotyping methods include the ability to generate micro-haplotypes by discovering additional mutations within the close proximity of targeted variants, which can greatly improve inclusion or exclusion of closely related individuals of the same species (parentage association). Another big advantage of the AgriSeq targeted GBS technology is the flexibility to modify the panel content and thus add or delete markers from those being used to interrogate targeted genomic loci by sequencing. The panel is simply a pool of PCR primers used in the highly-multiplexed target amplification reaction that is the first step in AgriSeq NGS library production. This pool of primers can be altered by mixing multiple pools together, adding additional primers to an existing pool, or by deleting markers through re-pooling of individual amplicon primer sets also provided with panel orders. Furthermore, the flexibility of the AgriSeq targeted GBS technology allows one to combine libraries from multiple species on a single chip run which improves the economics when one does not have enough samples from a single species to fill up a sequencing chip (see illustration in Figure 2 below).

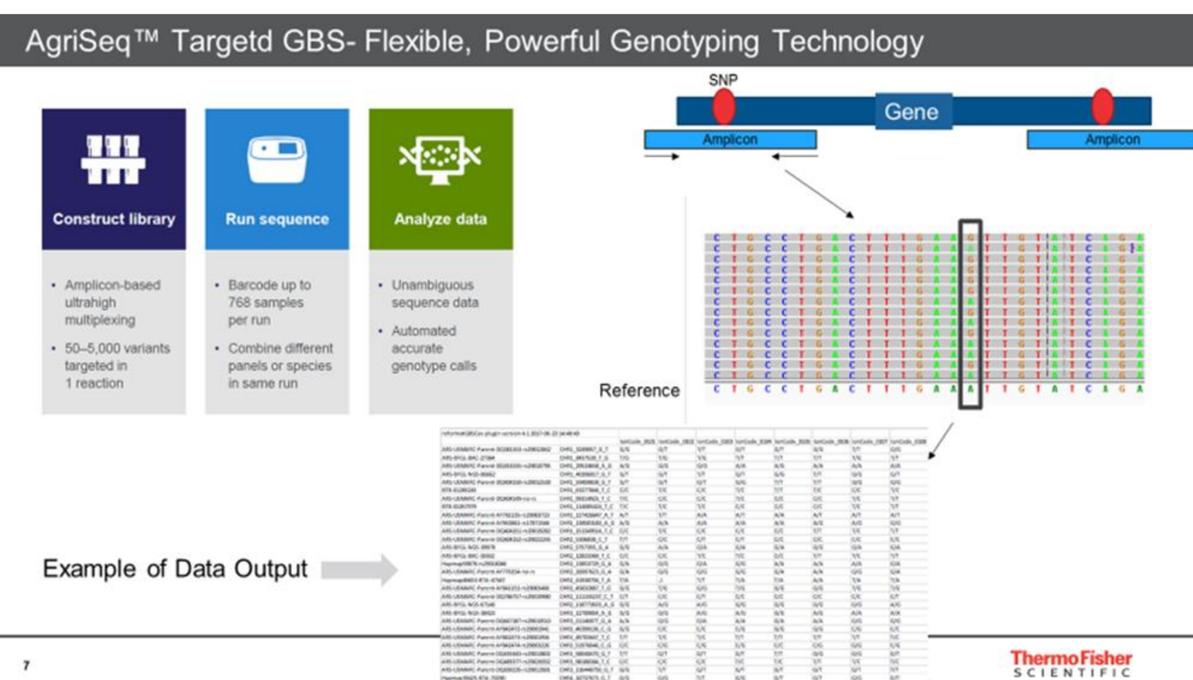


Figure 1. AgriSeq Targeted GBS Overview and Workflow. AgriSeq tGBS consists of 3 steps following panel design, 1] targeted amplicon library preparation, amplicon sequencing, and data analysis (sequence alignment and variant calling)

AgriSeq Target Enrichment Provides Workflow Flexibility

Panels can be pooled and/or markers can be added (spiked into existing panels) to meet customers' needs.

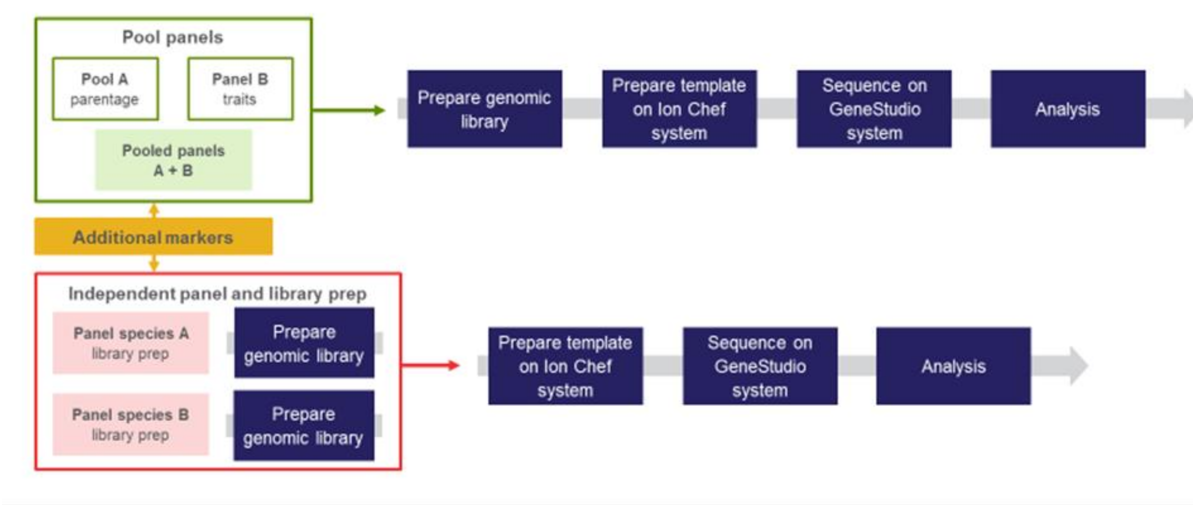
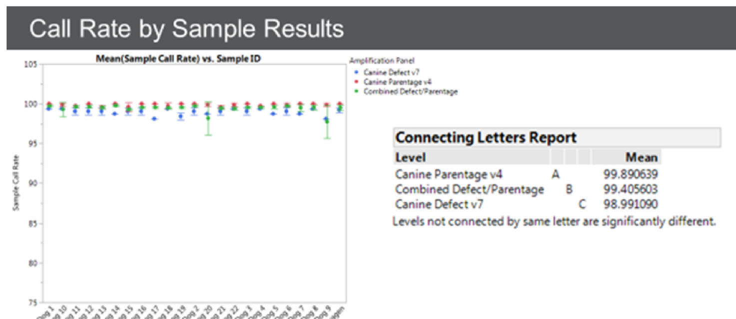


Figure 2. Illustration of panel and sample mixing during workflow

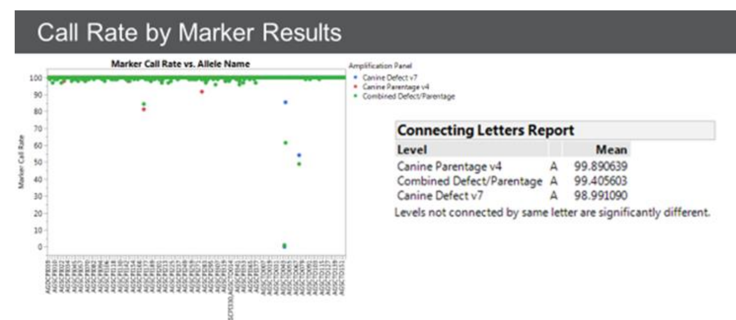
Results

- **Tested 24 samples in replicates (n=2 for separate panels and n=4 for combined LP) using the AgriSeq 96-well protocol.**
- **All libraries were pooled and sequenced on a 540 chip on the Ion S5XL.**
- **Analysis was performed with the files below:**

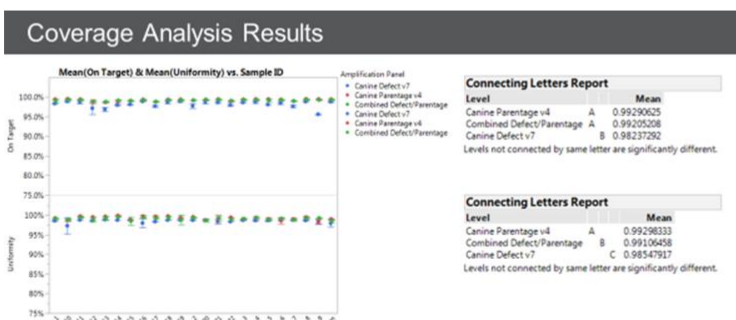
Reference genome:	ThreeAmigosCanineReference
Targeted regions:	AgriSeq_Canine_CTD_Region (Canine Defect)
	AgriSeq_Canine_CP_Region (Canine Parentage)
	AgriSeq_Canine_CTD_CP_Region (Combined LP)
Hotspot regions:	AgriSeq_Canine_CTD_Hotspot (Canine Defect)
	AgriSeq_Canine_CP_Hotspot (Canine Parentage)
	AgriSeq_Canine_CTD_CP_Hotspot (Combined LP)



- There was no drop-off in mean call rate by sample between the combined and separate library prep panels.



- There was no difference in mean call rate by marker between the three panels.



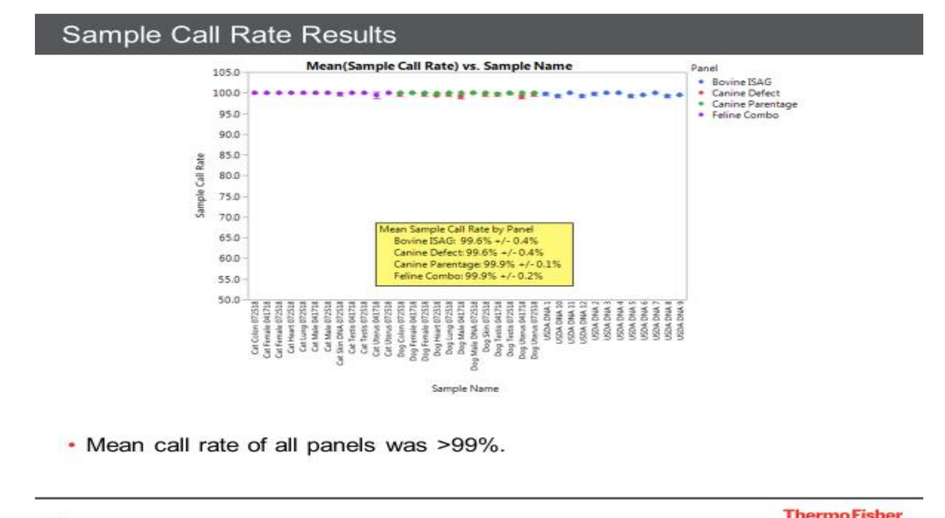
- While there were some small differences in % on target and uniformity between the panels, they were not statistically significant.

Figure 3A,B, &C. Data showing panel performance metrics do not change significantly when panels are run separately or after mixing them together. A. Sample call rate B. Marker Call rate. C. Coverage analysis are shown

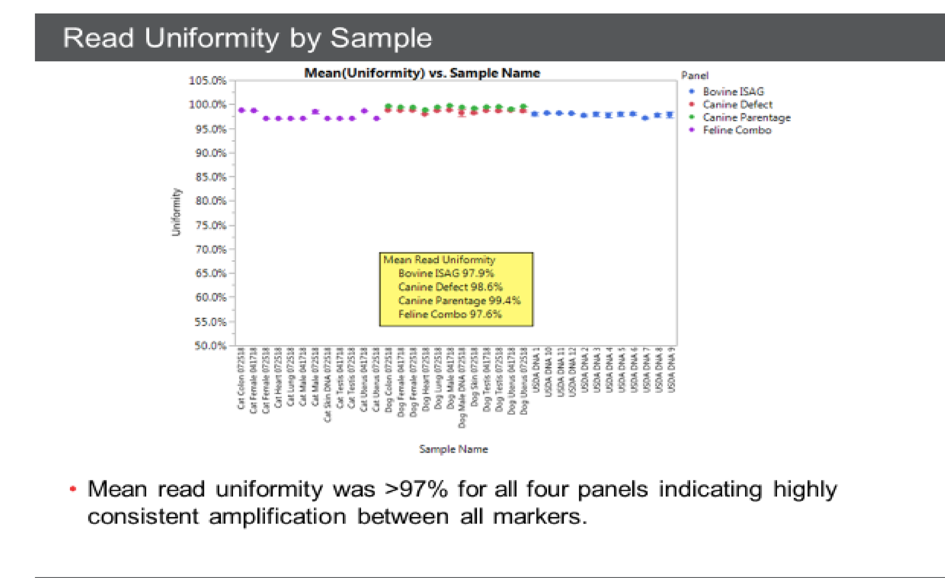
Results

Next We wanted to demonstrate that you can run libraries from different panels for the same species or even different species on the same sequencing run without impacting results in order to save on reagent costs and time.

- **4 different AgriSeq panels targeting three different species were chosen:**
  - **Canine Parentage; Canine Defect; Feline Combo; and Bovine ISAG panels were used to produce species-specific libraries**
  - **Each panel was run on 12 different DNA samples in replicate.**
- **Libraries were processed using the standard AgriSeq 96-well protocol.**
  - **18 amplification cycles were used which is standard for the Canine Defect, Feline Combo, and Bovine ISAG panel. This is one more amplification cycle than is recommended for Canine Parentage.**
- **Samples were combined and sequenced on the same Ion 540 chip run.**
- **When setting up the run plan, uncheck the box under the reference files that says "Use Same Reference & BED Files for All Chips"**
- **This allows you to choose individual analysis files for each sample.**
- **Initial mapping/alignment as well as all plugins use the individual analysis files so you don't have to go in and manually reanalyze.**
- **This feature avoid manual, sequential analysis, which is much more time consuming.**



- Mean call rate of all panels was >99%.



- Mean read uniformity was >97% for all four panels indicating highly consistent amplification between all markers.

Figure 4A & B Data showing panel performance metrics do not change significantly when libraries from different species are mixed in the same sequencing run. A. Sample Call Rate and B. Mean Read Uniformity by Sample are shown

CONCLUSION

- Same species panels can be run combined at will to alter panel content without impacting individual panel performance.
- Panels from different species can be run in the same sequencing reaction without impacting individual species panel performance which can save on sequencing reagent costs and sequencing time.
- Caveats:
  - Panels of significantly different sizes should be amplified on separate plates due to differences in cycle number.
  - Panels from different species should not be combined in the same post-ligation pool.
  - All samples run on the same plate should be normalized to the same concentration of DNA regardless of species.

For Research Use Only. Not for use in diagnostic procedures.

Thermo Fisher Scientific • 2130 Woodward Street • Austin, TX 78744 USA • thermofisher.com

©2018 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries.