Genetic marker screening for agrigenomics applications using Ion AmpliSeq technology

Key findings

- **Transformative**—a sequencing technology that is able to target known variants and discover novel variants through accurate and automated variant calling that helps eliminate bias when characterizing molecular markers.

- **Customizable**—Ion AmpliSeq™ technology, in combination with Ion AmpliSeq™ Designer (Figure 1), simplifies molecular marker assay design for agricultural genomics (agrigenomics) applications of any species through customizable panels that enable unbiased target choice that can encompass multiple variants per target region.

- **Flexible**—the simplicity of PCR and the ability to barcode hundreds of samples onto a single Ion Torrent™ sequencing chip results in a rapid 2 to 3 day workflow for the cost-effective processing of hundreds of samples and thousands of targets per week. Further, Ion AmpliSeq™ panel content is scalable with the flexibility to iteratively change content to accelerate marker selection and breeding programs using a single integrated workflow.

- **Concordant**—data obtained through collaborations resulted in single nucleotide variants identified across four different Ion AmpliSeq panels (targeting three different species) that were highly concordant (>95%) with bead-based microarray data.

Figure 1. The Ion AmpliSeq Designer workflow for molecular marker screening. Multiple options are possible, including designing your own (shown here) or receiving assistance from your field bioinformatics scientist or from our in-house Ion AmpliSeq panel specialists. DNA hotspot designs are optimized for unbiased target design that, unlike microarrays, can be designed to cover multiple variants (SNPs and short indels) per target region. Ion AmpliSeq technology and design expertise has enabled successful panel designs to be created for cow, rice, barley, tomato, salmon, wheat, sorghum, horse, dog, corn, *Brachypodium*, spruce, and oat.

ThermoFisher SCIENTIFIC

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

The advantages of NGS

In agrigenomics, next-generation sequencing (NGS) methods have advanced the discovery and selection of molecular markers used for the identification and screening of potentially complex traits. In contrast to microarray-based genotyping that requires a priori knowledge of variants present in a population for the targets to be present on the array, certain NGS-based genotyping methods such as restriction site–associated DNA sequencing (RAD-Seq) can be performed without a reference genome [1]. Likewise, NGS methods, such as Ion AmpliSeq™ workflows, can provide an unbiased targeted genotyping approach to interrogate a greater diversity of molecular markers including single nucleotide variants (SNVs), multiple nucleotide variants (MNVs), small insertions or deletions (indels), and short tandem repeats (STRs). Additionally, NGS genotyping facilitates characterization of highly diverse organisms, which can be challenging and potentially results in bias when microarray-based genotyping is used to investigate a population with a highly variable genetic background.

Ion AmpliSeq panels for marker screening—one technology, many species

While genome-wide approaches such as RAD-Seq are appropriate for marker discovery [1], a smaller number of molecular markers are typically selected for screening and validation in marker-assisted selection studies using smaller, focused marker panels and larger sample numbers. Filtered, high-quality variants identified by genome-wide discovery methods have flanking sequence information for use in the development of downstream Ion AmpliSeq panels that can be custom designed for any species of interest (Figure 1). By leveraging the multiplexing capability of Ion AmpliSeq technology, targeting of up to several thousand known user-defined molecular markers can be accomplished by PCR amplification.

Using a free online assay design tool, Ion AmpliSeq Designer, primer pools can be created to target any customer-defined region of interest within a reference sequence (Figure 1). For agrigenomic applications, the reference genomes for cow, pig, sheep, corn, rice, soybean, and tomato are preloaded into Ion AmpliSeq Designer (ampliseq.com). Additionally, Ion AmpliSeq Designer allows the creation of Ion AmpliSeq panels for private reference genomes or known target regions in a secure cloud computing environment.

Automatable workflow for rapid and cost-effective molecular marker screening

The simplicity of PCR-based targeting results in a rapid workflow for high-throughput sample processing (Figure 2). The Ion AmpliSeq workflow is partially automatable, facilitating the speed necessary to address time-sensitive projects and accelerate molecular-based breeding decisions.

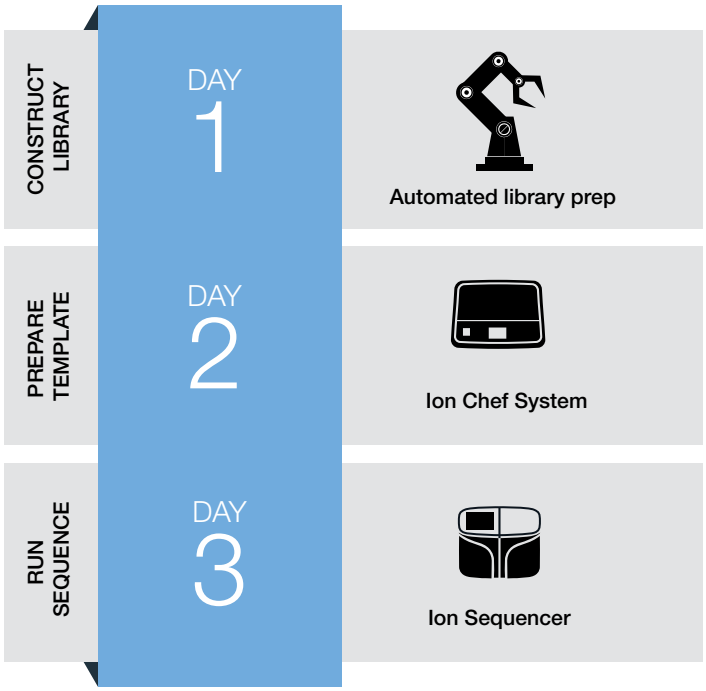

Figure 2. The automated Ion AmpliSeq workflow for molecular marker screening. Ion AmpliSeq library is constructed and barcoded using liquid-handling robotic systems (7.5 hours). Sequencing template is then prepared and chips are loaded using the Ion Chef System (13 hours). The sequencing reaction is run (2.5 hours) and data analyzed using Torrent Suite Software and Torrent Variant Caller. Note that Torrent Variant Caller is only compatible with diploid genomes.

Using liquid-handling robots, barcoded Ion AmpliSeq™ libraries can be constructed in 7.5 hours. This is followed by multiplexed template preparation overnight using the Ion Chef™ System (13 hours) and sequencing on the Ion S5/S5XL or the Ion Proton™ instrument (2.5 hours). Including signal processing and variant calling, the total workflow can be completed in 2 to 3 days for the processing of hundreds of samples with targeted genotyping of thousands of molecular markers.

Torrent Suite™ Software is optimized to de-multiplex barcoded samples from each Ion Torrent™ chip and automatically analyze the targeted regions for coverage and produce variant allele calls, resulting in high-quality genotyping data for the complete amplicon region flanking the variant of interest.

Results from the International Rice Research Institute (IRRI) rice Ion AmpliSeq™ panel

IRRI, located in Los Baños, Philippines, aims to reduce poverty and hunger, improve the health of rice farmers and consumers, and ensure environmental sustainability of rice farming. The IRRI Ion AmpliSeq™ panel targets 500 rice genetic markers across a total of 84,572 bp in a single primer pool (Table 1). The average call rate observed was >94% for the panel when tested on validation samples. Using 56 samples, the average genotype concordance was >98.9% for 299 positions in common between the IRRI Ion AmpliSeq panel and GoldenGate™ microarray data (Figure 3).

Using 46 samples, the average genotype concordance was >99% for 231 positions in common between the IRRI Ion AmpliSeq panel and Infinium™ microarray data. Further, using 5 rice samples, the average genotype concordance was >99.5% for 58 positions when the IRRI Ion AmpliSeq panel results were confirmed using Sanger sequencing.

Table 1. Ion AmpliSeq™ molecular marker screening panel characteristics for three different agriculturally important species.

Customer	Organism	No. of markers	Total targeted base pairs	No. of samples	Concordance*
GeneSeek, Inc.	Cow	4,802	556,881	239	>95%
Delta Genomics	Cow	121	14,153	160	>98%
IRRI	Rice	500	84,572	46	>99%
Floragenex	Barley	320	59,840	93	>99%

*Calculated from subset that overlaps historical data.

Table 1. Ion AmpliSeq™ molecular marker screening panel characteristics for three different agriculturally important species. Replicate barcoded libraries for GeneSeek, International Rice Research Institute (IRRI), and Delta Genomics samples were prepared using standard Ion AmpliSeq™ 2.0 reagents and protocols, using the Tecan™ Freedom EVO™ NGS robotic liquid handling system. The Ion Library Equalizer™ Kit was used for on-robot normalization of libraries prior to sample pooling. The Ion PI™ IC 200 Kit was used for template preparation and chip loading on the Ion Chef System. Sequencing was performed on the Ion Proton instrument using the Ion PI™ Chip Kit v2. Replicate barcoded libraries for Floragenex samples were prepared using standard manual Ion AmpliSeq™ 2.0 reagents and protocols. The Ion PGM™ Template OT2 200 Kit was used for template preparation on the Ion OneTouch™ 2 System. Sequencing was performed on the Ion PGM™ Sequencer using the Ion PGM™ Sequencing 200 Kit v2.

The GeneSeek experience with Ion AmpliSeq technology

GeneSeek, part of Neogen Corporation, is a leading provider of genetic testing solutions to the cattle industry. Stewart Bauck, general manager at GeneSeek/Neogen Agrigenomics, describes implementing Ion AmpliSeq technology into their product offerings: “For us the transformative event is the impact that it has on our workflow processes and the potential impact on cost. We can change the target list very frequently. We can modify it. We can evolve it rapidly. For us those are all the things that are really important.”

The Ion AmpliSeq panel designed by GeneSeek for bovine screening targets 4,802 markers across a total of 556,881 bp in a single primer pool (Table 1). Using 239 cattle samples, the average genotype concordance was >95% for 3,038 positions in common between the sequencing and bead-based microarray results (Figure 3). Dr. Bauck describes the optimal balance that Ion AmpliSeq technology fulfills for Geneseek’s genotyping needs: “We consider the real sweet spot for the market to be that 1,000 to 5,000 SNPs can be targeted in a single Ion AmpliSeq panel, which for us is a huge part of our business. This is an area where we can optimize our throughput and processes to take advantage of the technology and improve greatly our efficiency without sacrificing anything in terms of data quality.”

The Ion AmpliSeq panels designed by GeneSeek have also enabled new low-density genotyping solutions. Dr. Bauck explains, “People have spent significant time and money genotyping important animals with high-density arrays. Now they want to move downstream and leverage that information by genotyping a smaller SNP set for a more efficient and more cost-effective solution. This technology, with the ability to pool large numbers of samples and really leverage the capacity of Ion AmpliSeq panels are things that make this attractive for us.”

ion torrent

Results from the Delta Genomics bovine Ion AmpliSeq™ panel

Delta Genomics is a national, not-for-profit provider of genomics services including genotyping, sequencing, and biobanking for the livestock industry and research community. The Delta Genomics cattle parentage Ion AmpliSeq™ panel targets 121 markers across a total of 14,153 bp in a single primer pool (Table 1). The average call rate was >95% and the average genotype concordance was >98% for 104 positions in a comparison between the Delta Genomics Ion AmpliSeq™ panel and bead-based microarray results for 160 cattle samples (Figure 3).

Results from the Floragenex barley Ion AmpliSeq panel

A leader in RAD-Seq, Floragenex provides genomic services for discovery, genotyping, and analytics for any species. The Floragenex barley Ion AmpliSeq™ panel targets 320 markers across a total of 59,840 bp in a single primer pool (Table 1). Of the 320 markers, 112 were expressed sequence tag (EST) derived and genotyped by BeadXpress™ microarray while the remaining 208 markers were genotyped by RAD-Seq experiments. With an average call rate of >98%, the average genotype concordance for 93 barley samples was >99% for 96 positions in common between the sequencing and bead-based microarray approaches (Figure 3).

Conclusions

Scalable Ion Torrent™ instrumentation and chips, in combination with Ion AmpliSeq technology, allow molecular marker discovery and screening on a single platform for the genotyping of any organism, including highly diverse organisms that can be a challenge for microarray-based approaches.

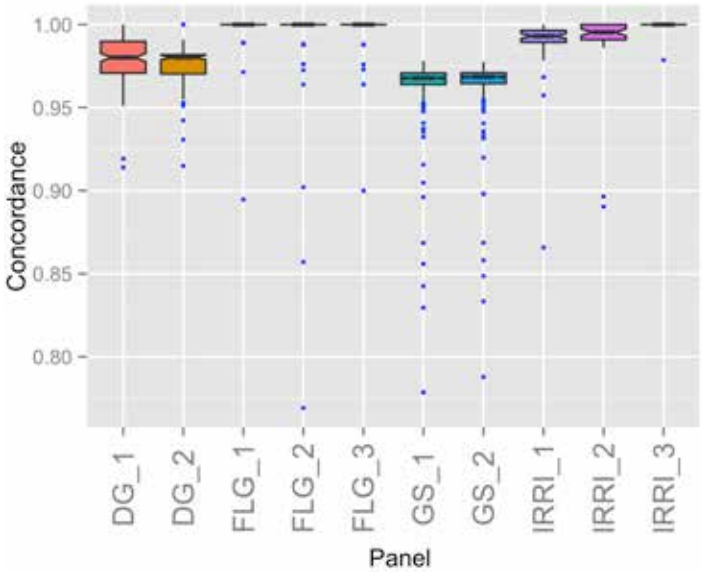

Figure 3. Excellent genotype concordance between Ion AmpliSeq™ molecular marker panels and orthogonal genotyping analyses. Concordances between different genotyping assays (excluding no-calls by both assays) for replicate runs are shown for each panel (DG, Delta Genomics; FLG, Floragenex; GS, GeneSeek; IRRI, International Rice Research Institute). IRRI_1 is in concordance with GoldenGate microarray data, IRRI_2 is in concordance with Infinium microarray data, and IRRI_3 is in concordance with Sanger sequencing data.

Reference

1. Davey JW, Hohenlohe PA, Etter PD et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12:499–510.

Find out more at **thermofisher.com/gbs**

ThermoFisher SCIENTIFIC

For Research Use Only. Not for use in diagnostic procedures. © 2015 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. **CO210586 1115**