applied biosystems

High quality consistent genotypes using Applied Biosystems™ Eureka™ Genotyping panels of varying sizes

Victor V. Missirian[1], John D. Curry[1], Ali Pirani[1]
[1]Thermo Fisher Scientific, 3450 Central Expressway, Santa Clara, California, USA, 95051

INTRODUCTION

Applied Biosystems™ Eureka™ Genotyping Solution is a low cost, high-throughput targeted genotyping by sequencing platform that supports the detection of tens to thousands of genetic markers (SNPs and insertions/deletions). It has been successfully used for a variety of applications (parentage, sex validation, genomic evaluation, carrier diseases) both in crops and animals.

After sample processing, Eureka™ next generation sequencing (NGS) read counts are obtained for each sample, at each genetic marker. These counts are appropriately scaled, normalized, and transformed, before calling genotypes in a cluster-based Bayesian framework (BRLMM-P). The process can rapidly and reliably enable high quality genotyping results across a wide range of panel sizes.

We demonstrate on four Eureka panels of varying size—two in the low range (tens - 500 markers) and two in the high range (500 - 3000 markers)—that as panel size increases, we continue to obtain high sample pass rate, call rate, and concordance to known genotypes.

METHODS

Eureka™ genotyping assay

Eureka Genotyping Solution utilizes NGS to enable genotyping of thousands of DNA samples for tens to thousands of markers. The Eureka genotyping assay is a ligation-dependent PCR reaction that uses interrogation site bar codes contained within the ligation probes as well as sample index bar codes added during the amplification step. NGS libraries can be created for thousands of DNA samples within 24 hours. Short-cycle sequence data is generated from the prepared libraries, and software is used to tabulate the number of reads that contain each combination of sample, locus, and allele bar code (as appropriate). The genotype of each sample for each locus is inferred from statistical analysis of the tabulated reads.

Genotyping analysis

Sample processing returns Eureka™ NGS read counts for each sample, at each genetic marker. Counts are normalized, scaled, and then transformed to expand the central call region according to a stretch constant K. Genotypes are then called in a cluster-based Bayesian framework (BRLMM-P) that adapts pre-positioned genotype cluster locations called “priors” to the sample data and computes three posterior cluster locations.

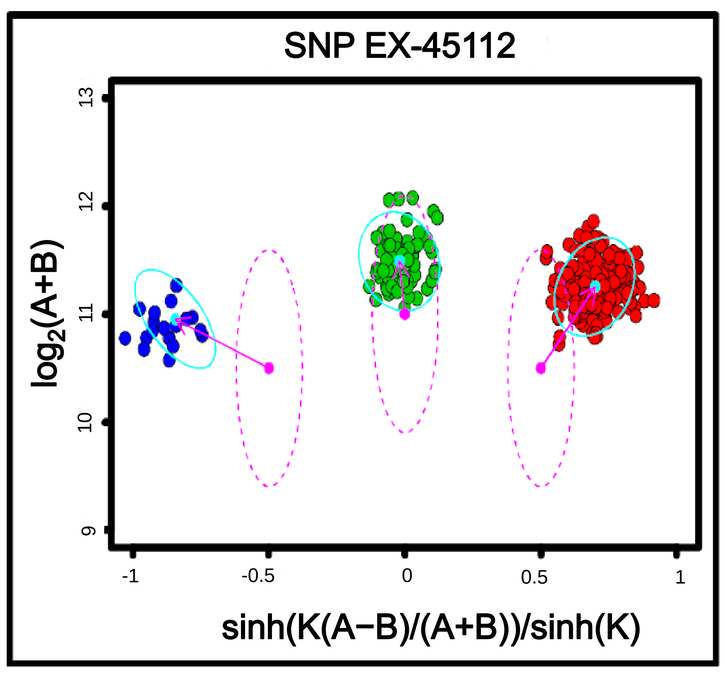

Figure 1. Example of a SNP cluster plot.
Eureka™ panels contain tens to thousands of markers. Performance of each can be visualized with a cluster plot. Each sample is represented by a single point, whose coordinates reflect its (scaled, normalized, transformed) read counts. Prior location (dashed oval) and observed samples (points) are then combined to get a posterior (solid oval), which is used to call genotypes.

Testing

Eureka™ genotyping was evaluated on four panels of different size: two in the low range (tens - 500 markers) and two in the high range (500 - 3000 markers). For each panel, we individually genotyped a single set of samples. Performance on each set of samples was measured in four ways: sample pass rate, call rate, concordance to known genotypes, and visual evaluation of cluster plots.

Sample pass rate: Samples are considered not passing if they have a mean call rate (see below) of less than 90%. Reported values of the next three measures were computed over passing samples.

Call rate: Genotyping uses a cluster-based Bayesian framework to compute a most likely genotype (AA, AB, or BB) and a confidence score that reflects the probability of that genotype. A (marker, sample) pair is called only if the confidence score of the most likely genotype exceeds a specific threshold, otherwise it is no-called (assigned no genotype). Call rate is the percentage of pairs that are called.

Concordance: Eureka™ genotyping calls (AA, AB, or BB) can be compared to reference calls from a second source. Concordance is the percent of (marker, sample) pairs with calls from both Eureka™ and reference for which the Eureka™ call matches the reference call.

Cluster plots: We performed visual evaluation of selected cluster plots, taking into account clarity and cluster separation.

RESULTS

Performance

One way to measure performance is by computing sample pass rate, call rate, and concordance, as defined in the previous section. The table below displays sample pass rate for each panel as well as mean call rate and concordance across all markers. These measures are consistently high across all tested panels.

Panel	Number of markers	Sample pass rate	Call rate (mean)	Concordance (mean)
Bovine #1	122	98.9%	99.2%	99.7%
Diploid plant, grain	418	96.9%	98.7%	99.7%
Bovine #2	515	95.8%	98.9%	99.8%
Diploid plant, legume	977	100.0%	99.3%	99.5%

Table 1. Sample pass rate, call rate, and concordance.
Performance of a Eureka™ panel can be measured in terms of sample pass rate, call rate, and concordance. We have calculated these measures on four panels of different size.

Visual Evaluation

A second way to measure performance is the visual evaluation of cluster plots. A high performance SNP is expected to have a clear and well separated cluster pattern. The figure below displays cluster plots of three example SNPs (column) from each tested panel (row). Cluster patterns are clear and well resolved across all tested panels.

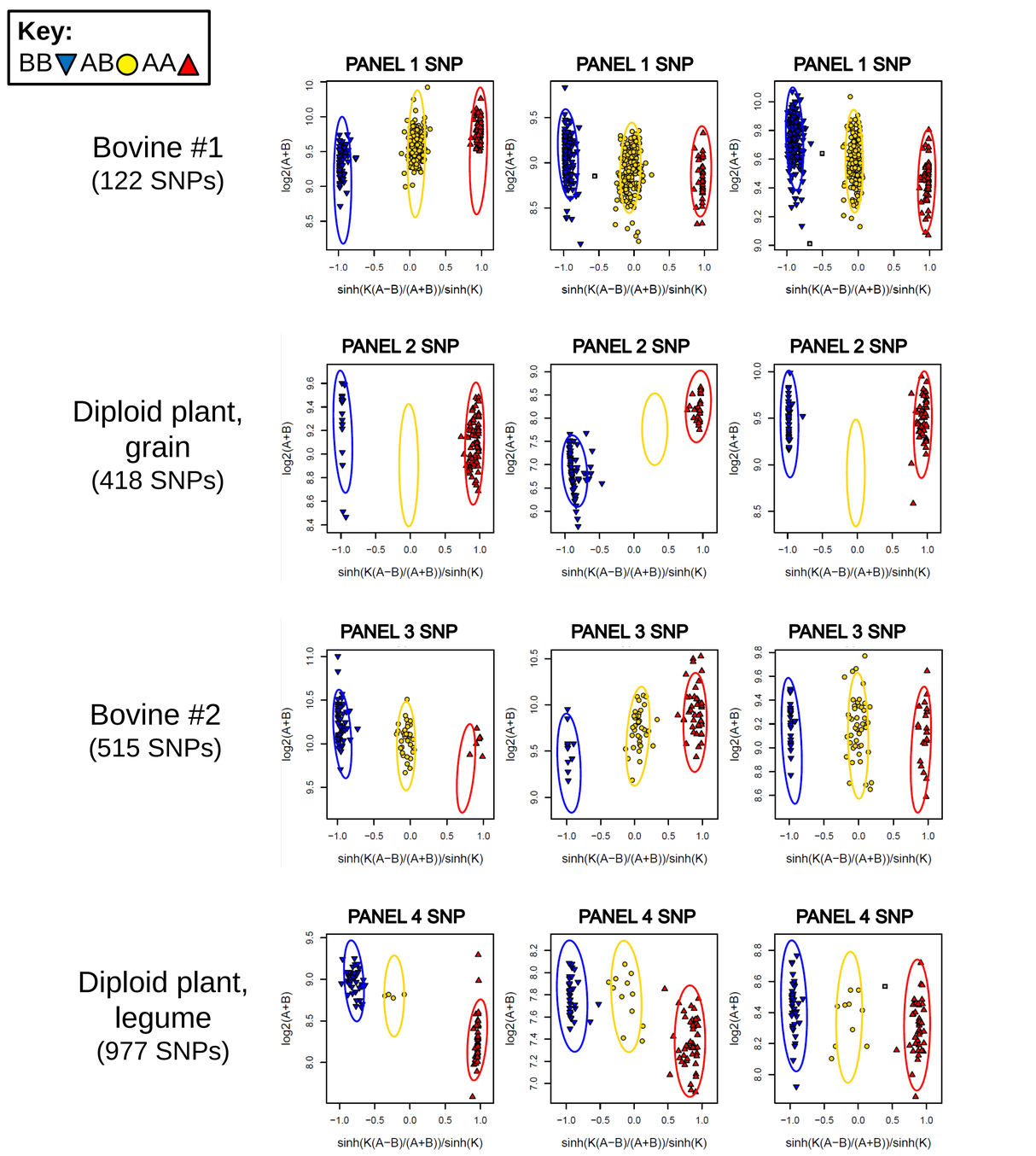

Figure 2. Cluster plots of example markers.
Performance of a Eureka™ panel can be measured by the clarity and separation of cluster plots for each marker. We have cluster plots on three example markers from each of the four tested panels. Colors and shapes of points correspond to the three genotyping calls (AA, AB, or BB).

CONCLUSIONS

Eureka™ genotyping has consistently high performance across a wide range of panel sizes, as shown by sample pass rate, call rate, and concordance as well as by clear and well separated cluster plots. Eureka™ genotyping is currently qualified for panel sizes of tens of markers to 3000 markers.

ThermoFisher SCIENTIFIC

For Research Use Only. Not for use in diagnostic procedures.
© 2019 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.